


```
QY 385 GGGCCCAAGTCAGAGAGTGGGTAGGGTCTCCAGACACAGCCCTCCCTCTGGGGCAA 444
DB 3078 TCCGAGTCAGATTGATGTGGAAAGGTGTACACAGCAAAAGCCCCCTCCCAATCGGGGCCA 3137
QY 445 CATGCTCTGCTCTGAGAGCTTGGCCACAGCTCCTGTCTCATTTGAGAGCTGC 493
DB 3138 CATGCTCAGCTC--AGGACTGGGGCCCTGTCTCTTCTTGGGTGAGTCTGC 3184
```

RESULT 2

```
US-10-750-623-50617
; Sequence 50617, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50617
; LENGTH: 3396
; TYPE: DNA
; ORGANISM: Bovine 1986680611167
; US-10-750-623-50617
```

```
Query Match 28.8%; Score 142.2; DB 8; Length 3396;
Best Local Similarity 73.0%; Pred. No. 5.8e-28;
Matches 211; Conservative 0; Mismatches 73; Indels 5; Gaps 2;
```

```
QY 208 GTCTGCTGAGCTCATCAGAACTGGCATCTGCAGAGATG---GAAGTGAATTCGGGGA 264
DB 2898 GACTTTCAGCCCTCATCAGAAACCGGAGCTGCAGAGATGACAGAAAGTGAATTCAGGGA 2957
QY 265 AGAAAAATGTTTTCACCAACAATCCCAAAGTCAGAGATTTCTCATCATCCCATGATC 324
DB 2958 AGAAATGATTTTTCACCAACAATCCCAAAGACTGGGTTTCTCATCTTCCACGAGATC 3017
QY 325 CTATTTCCCGAGCATGCTGCAAGCGGTAGTCTGTCTAGCAAGGGCTGTGGAGAAAGGGCCA 384
DB 3018 TTAATTCCTCCGCAATCCCGTCAAGTATCATCTCATAGGCTTCTGAGAGAGGGCCAG 3077
QY 385 GGGCCCAAGTCAGAGAGTGGGTAGGGTCTCCAGACAGGCCCTCCCTGTGGGGCAA 444
DB 3078 TCCGAGTCAGATTGATGTGGAAAGGTGTACACAGCAAAAGCCCCCTCCCAATCGGGGCCA 3137
QY 445 CATGCTCTGCTCTGAGAGCTTGGCCACAGCTCCTGTCTCATTTGAGAGCTGC 493
DB 3138 CATGCTCAGCTC--AGGACTGGGGCCCTGTCTCTTCTTGGGTGAGTCTGC 3184
```

RESULT 3

```
US-11-036-196-1780
; Sequence 1780, Application US/11036196
; Publication No. US20060078900A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Higgs, Brandon
; APPLICANT: Castle, Arthur
; APPLICANT: Elishoff, Michael
; TITLE OF INVENTION: Molecular Toxicology Modeling
```

```
; FILE REFERENCE: 44921-5089-US
; CURRENT APPLICATION NUMBER: US/11/036,196
; CURRENT FILING DATE: 2005-01-18
; PRIOR APPLICATION NUMBER: US/10/152,319
; PRIOR FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/292,335
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/297,523
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,925
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,810
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,807
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,808
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/315,047
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/324,928
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/330,867
; PRIOR FILING DATE: 2001-11-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1780
; LENGTH: 15231
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. NM_022401
; US-11-036-196-1780
```

```
Query Match 10.2%; Score 50.2; DB 11; Length 15231;
Best Local Similarity 49.8%; Pred. No. 0.0026;
Matches 127; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
```

```
QY 20 CGAGTCCCAAGAGCCCGAATTTGAGAGCTGTGTGCAGAAATTCCTCCGAGAGGCCCA 79
DB 7947 CGAGAGGAGGAAGGCGCAAGCTGTGAGCAGCTTTCCAGACGAGGTGCAAAAGCAAAACA 8006
QY 80 GCTGTCCCTGCTGAGGAAGCGCTGCTGTGATGAGAAATCTTCTCCGAGAGTGC 139
DB 8007 GCTTCAGAGAGCGCAGCAGCGGACAGCAGCAGATGAGCAGGAAAAAGCAGAGCTGT 8066
QY 140 TCCCGTGTGTCAGAGAGAGCTCAGAGAGCTGTGACAGATGTGTGGCGGCTTGTAGGCTGT 199
DB 8067 GGCACAGATGAGAGAGGCGCCGAGGCGGAGCGTGAAGCAGAGAGAGGAGGTGTAGGCGCA 8126
QY 200 GGAAGAAAGTCTGTGAGAGCTCATCAGAAATCTGCAATCTGTCAGAGAGATGGAAGTGAATTC 259
DB 8127 GCAAGAGAAATCTGAGAGCTGTGTGAGACAGCAGCGGAGAGAGAGAAACTTACTGGCAGA 8186
QY 260 GGGAGAAAGAAATGT 274
DB 8187 GGAAGACAGAGGCT 8201
```

RESULT 4

```
US-11-136-527-306
; Sequence 306, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
```

SOFTWARE: PatentIn version 3.2
 SEQ ID NO 306
 LENGTH: 15234
 TYPE: DNA
 ORGANISM: Rattus norvegicus
 US-11-136-527-306

Query Match
 Best Local Similarity 49.8%; Score 50.2; DB 14; Length 15234;
 Pred. No. 0.0026;
 Matches 127; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

20 CGAGTCCCAAGAGCCGAGTTTGAAGGCTGTGGCAGAAATTCGCCGAGAGAGGCCCA 79
 7947 CGAGCAGAGAGAGCCCAAGCTTGAGCAGCTTTTCCAGAGCAGAGTGGCAAAACCA 8006
 80 GCTGTCCCTGTGAGAGCCGAGGCTGTGGTGTATGAGAGATCTTCTCCGAGAGGTGC 139
 8007 GCTGCAAGAGAGCAGCAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 8066
 140 TGCCTGTGTGACAGAGAGCTCAGAGAGCTGCGAGAGTCTGTGGCCCTTGAAGCTGCT 199
 8067 GGCACAGATGAGAGAGCCCGAGAGCGCGCAGCGTGAAGCAGAGAGGCTGTGAGGCCAA 8126
 200 GGAAGAAAGTCTGTGAGCTCATCAGAACTGGCATCTGCAGAGATGAGAGTGAATTC 259
 8127 GCAAGAGAACTGACGCTGTGAGCAGCAGCGCAGCAGCAGCAGCAGCAGCAGCAGCAG 8186
 260 GGGGAAAGAAATGCT 274
 8187 GGAGAACCAAGAGCT 8201

RESULT 5
 US-11-121-086-5
 ; Sequence 5, Application US/11121086
 ; Publication No. US20050266459A1
 ; GENERAL INFORMATION:
 ; APPLICANT: POULSEN, TIM S.
 ; APPLICANT: NIELSEN, KIRSTEN V.
 ; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
 ; FILE REFERENCE: 09138, 6000-00000
 ; CURRENT APPLICATION NUMBER: US/11/121,086
 ; CURRENT FILING DATE: 2005-05-04
 ; PRIOR APPLICATION NUMBER: 60/567,570
 ; PRIOR FILING DATE: 2004-05-04
 ; NUMBER OF SEQ ID NOS: 107
 ; SOFTWARE: PatentIn version 3.3
 ; SEQ ID NO 5
 ; LENGTH: 153376
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-11-121-086-5

Query Match
 Best Local Similarity 53.4%; Score 48.2; DB 14; Length 153376;
 Pred. No. 0.014;
 Matches 101; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

2 GGGCGGGCGGGGAGATGCCAGTCCCAAGAGCCGAGTTTGAAGGCTGTGGCAGAAAT 61
 16201 GGAGGAAACAG 16260
 62 CCCGAGAGAGAGAGCCAGCTGTCTCCCTGTGAAAGCCAGAGGCTGTGTGTATGAGAA 121
 16261 ACAG 16320
 122 GTCTTCTCCGAGAGGTGCTGCGGTGTGACAGAGAGCTCAGGAGCTGTCAGAGTCTGTG 181
 16321 GAGAGAACAG 16380
 182 GCGGCGCTT 190
 16381 GTGGGTCTT 16389

RESULT 6
 US-10-330-773-631/c
 ; Sequence 631, Application US/10330773
 ; Publication No. US20060040262A1
 ; GENERAL INFORMATION:
 ; APPLICANT: David W. Morris
 ; APPLICANT: Marc Malandro
 ; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
 ; FILE REFERENCE: 529452001300
 ; CURRENT APPLICATION NUMBER: US/10/330,773
 ; CURRENT FILING DATE: 2002-12-27
 ; NUMBER OF SEQ ID NOS: 981
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 631
 ; LENGTH: 30191
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(30191)
 ; OTHER INFORMATION: n = A,T,C or G
 US-10-330-773-631

Query Match
 Best Local Similarity 52.3%; Score 46.6; DB 9; Length 30191;
 Pred. No. 0.027;
 Matches 103; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

9 CCGGGGATGCCAGTCCCAAGAGCCGAGTTTGAAGGCTGTGGCAGAAATTCGCCGAG 68
 13754 CAGAGCAG 13695
 69 AAGAGGCCCAAGCTGTCTCCCTGTGTGAGAGCCAGAGGCTGTGTATGAGAAATCTTCT 128
 13694 CAG 13635
 129 CCGGAGGCTGTCTCCCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 188
 13634 CAGAGCAG 13575
 189 TTGAGGCTGTGAGAG 205
 13574 CAGAGCTGAGAGAGAG 13558

RESULT 7
 US-10-330-773-905
 ; Sequence 905, Application US/10330773
 ; Publication No. US20060040262A1
 ; GENERAL INFORMATION:
 ; APPLICANT: David W. Morris
 ; APPLICANT: Marc Malandro
 ; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
 ; FILE REFERENCE: 529452001300
 ; CURRENT APPLICATION NUMBER: US/10/330,773
 ; CURRENT FILING DATE: 2002-12-27
 ; NUMBER OF SEQ ID NOS: 981
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 905
 ; LENGTH: 138808
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(138808)
 ; OTHER INFORMATION: n = A,T,C or G
 US-10-330-773-905

Query Match
 Best Local Similarity 52.7%; Score 45.2; DB 9; Length 138808;
 Pred. No. 0.086;
 Matches 96; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

30 GAGGCCAGTTTGAAGGCTGTGTGACAGAAATTCGCCGAGAGAGAGCCAGCTGTCTCTG 89


```
; CURRENT APPLICATION NUMBER: US/11/036,196
; CURRENT FILING DATE: 2005-01-18
; PRIOR APPLICATION NUMBER: US/10/152,319
; PRIOR FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/292,335
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/297,523
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,925
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,810
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,807
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,808
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/315,047
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/324,928
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/330,867
; PRIOR FILING DATE: 2001-11-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2106
; LENGTH: 648
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. U02506
US-11-036-196-2106
```

```
Query Match      8.5%; Score 42; DB 11; Length 648;
Best Local Similarity 47.7%; Pred. No. 0.22;
Matches 123; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
```

```
QY 47 GCTGTGGGAGAAATTCGCCGAGAGAGGCCCGCTGTCCTCCGTGTCAGAGCGGAGGCTG 106
DB 95 GTTGGTTTCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 154
QY 107 GCTGTGATGAGAAATTCCTCCGAGAGGCTGCTCCGTGTCAGAGAGAGCTCAGGGA 166
DB 155 GAGAGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 214
QY 167 GCTGCGAGAGTGTGGCGGCTTGAAGCTGTGAGAAAGTCTGAGCCTCATCAG 226
DB 215 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 274
QY 227 AAACGTGCATCTCCAGAGATGGAATGATTCGGGGAGAAATGTTTCACCAACAA 286
DB 275 GAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 334
QY 287 CATCCCAAGTCAGATT 304
DB 335 GAAAGAAAGAGAAATTT 352
```

```
RESULT 15
US-11-096-568A-21841/C
; Sequence 21841, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 21841
; LENGTH: 1811
; TYPE: DNA
```

```
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1811)
; OTHER INFORMATION: Ceres Seq. ID no. 12406472
US-11-096-568A-21841
```

```
Query Match      8.5%; Score 41.8; DB 11; Length 1811;
Best Local Similarity 50.8%; Pred. No. 0.3;
Matches 100; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
```

```
QY 2 GCGCGCCCGGGGAGATGCCGAGTCCCAAGAGCCGAGTTTGAGAGCTGTGGCAGAAAT 61
DB 573 GGCCTCTCTGACGGGTGCCGAGATCCATGGGCGCTTGATGACGSCATGTGTGTGA 514
QY 62 CCGGAGAAAGAGAGCCCAAGCTGTCCCTGTGTGAAGCCGAGGCTGCTGTGATGAGAA 121
DB 513 GCGCGAGGCGCTCCACTCGACGGGCGGCTTGAAACACACGCTGCTGTGTCTTTCGCA 454
QY 122 GTCCTTCGAGAGGCTGCTCCGCTGTGTGAGAGAGAGCTCAGGAGCTGGCAGATCGTG 181
DB 453 GCTTCTGAGGATCTGCGCGCACCGCTTCCGATGCCCGGAGCACCGCGCGCGCG 394
QY 182 GCGGCTCTTGAAGCTGC 198
DB 393 GCGGCGCTGCTGCTGC 377
```

```
Search completed: April 27, 2006, 05:41:19
Job time : 3289 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 03:03:54 ; Search time 3013 Seconds
(without alignments)
9300.970 Million cell updates/sec

Title: US-10-071-510A-16

Sequence: 1 cggccgagccgggagatgcc.....ccctctcattgagcctgc 493

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapexc 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: gb_Da:*
2: gb_in:*
3: gb_env:*
4: gb_com:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_to:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hcg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	490.4	99.5	5502	8 HSM808450	BX648302 Homo sapi
2	347.6	70.5	1938	8 AK098471	AK098471 Homo sapi
3	347.6	70.5	3978	6 CO735548	CO735548 Sequence
4	345.4	70.1	3003	6 CO845922	CO845922 Sequence
5	345.4	70.1	3003	8 AK131436	AK131436 Homo sapi
6	283.4	57.5	150350	8 CNS01DMV	AL138539 Homo sapi
7	283.4	57.5	150350	14 CNS01DRC	AL117186 Homo sapi
8	212.2	43.0	3370	9 BC099694	BC099694 Mus muscu
9	140.6	28.5	165217	14 AC152545	AC152545 Bos tauru
10	121.4	24.6	176510	9 AC147375	AC147375 Mus muscu
11	57	11.6	110000	15 AP008207_101	Continuation (102
12	57	11.6	142268	15 AP003018	AP003018 Oryza sat
13	57	11.6	154248	15 AP003631	AP003631 Oryza sat
14	57	11.6	159749	15 AP003020	AP003020 Oryza sat
15	56.8	11.5	2279	2 AY571308	AY571308 Kukulcani
16	56	11.4	110000	15 AP008207_037	Continuation (38 o
17	56	11.4	154137	15 AP003215	AP003215 Oryza sat
18	55.4	11.2	110000	15 AP008214_114	Continuation (115

C 19	55.4	11.2	135511	15 AP005097	AP005097 Oryza sat
C 20	55.4	11.2	154441	15 AP005383	AP005383 Oryza sat
C 21	55.2	11.2	194736	9 AC126271	AC126271 Mus muscu
C 22	54	11.0	109894	14 AP005975	AP005975 Oryza sat
C 23	54	11.0	110000	15 AP008213_014	Continuation (15 o
C 24	54	11.0	132526	15 AP006451	AP006451 Oryza sat
C 25	54	11.0	147782	15 AP005604	AP005604 Oryza sat
C 26	53.8	10.9	55528	2 AC105764	AC105764 Caenorhab
C 27	53.8	10.9	213476	14 AC114589	AC114589 Mus muscu
C 28	53.6	10.9	78835	15 AC120528	AC120528 Oryza sat
C 29	53.6	10.9	110000	15 AP008217_098	Continuation (99 o
C 30	53.4	10.8	110000	15 AP008213_013	Continuation (14 o
C 31	53.4	10.8	169511	15 AP005309	AP005309 Oryza sat
C 32	53.4	10.8	211542	15 AC126275	AC126275 Mus muscu
C 33	53	10.8	541	15 CR354530	CR354530 Plums pin
C 34	53	10.8	110000	15 AP008207_416	Continuation (417
C 35	53	10.8	110000	15 AP008208_227	Continuation (228
C 36	53	10.8	128223	15 AP005643	AP005643 Oryza sat
C 37	53	10.8	176530	15 AP003259	AP003259 Oryza sat
C 38	52.8	10.7	125020	8 AF429315	AF429315 Homo sapi
C 39	52.8	10.7	217522	9 AC117639	AC117639 Mus muscu
C 40	52.8	10.7	227897	14 AC111962	AC111962 Rattus no
C 41	52.8	10.7	237997	14 AC093989	AC093989 Rattus no
C 42	52.6	10.7	253504	9 AC102602	AC102602 Mus muscu
C 43	52.2	10.6	97839	14 AC141520	AC141520 Rattus no
C 44	52.2	10.6	110000	15 AP008217_250	Continuation (251
C 45	52.2	10.6	110000	15 AP008208_094	Continuation (95 o

ALIGNMENTS

RESULT 1
HSM808450
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo;
1 (bases 1 to 5502)
Boecker,H., Boecker,M., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
Fobo,G., Han,M. and Wiemann,S.
The German Human CDNA Consortium
Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the CDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp686M07125) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcentzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cdna/.

COMMENT

1..5502
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686M07125"
/cissue_type="human uterus"
/clone_id="686 (synonym: nlcc3). Vector pSPORT1_Sfi; host
DH10B; sites SfiI + SfiIB
/dev_stage="adult"
5474

FEATURES

source
polysite
ORIGIN

Query Match 99.5%; Score 490.4; DB 8; Length 5502;
Best Local Similarity 99.8%; Pred. No. 1,4e-96;
Matches 491; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGCGGCGCGGGGAGATGCCGAGTCCCAAGAGCGGAGTTTGAGAGGCTGTGGCAGAAATT 61
DB 3918 GGGGGGCCCCGGGGGATGCCAGTCCCAAGAGCGGAGTTTGAGAGGCTGTGGCAGAAATT 3977

QY 62 CCGGAGAAAGAGAGCCCAAGCTGTCTCTGTGTGAAGCGCAGGCGCTGGTGTATGAGAA 121
DB 3978 CCGGAGAAAGAGAGCCCAAGCTGTCTCTGTGTGAAGCGCAGGCGCTGGTGTATGAGAA 4037

QY 122 GTTTTCTCCGAGAGGCTGTCCGTTGTCAGAGAGAGCTGAGGAGCTGGCAGAGTCG 181
DB 4038 GTTTTCTCCGAGAGGCTGTCCGTTGTCAGAGAGAGCTGAGGAGCTGGCAGAGTCG 4097

QY 182 GCGGCGCTTGAGGCTGTGGAAGAAAGTCTGTGAGCCTCATGAGAACTGGCATCTGCA 241
DB 4098 GCGGCGCTTGAGGCTGTGGAAGAAAGTCTGTGAGCCTCATGAGAACTGGCATCTGCA 4157

QY 242 GAGAGTGAAGTGGATTTGGGGAGAAATGTTTTTACCACACATCCCAAAATCAG 301
DB 4158 GAGAGTGAAGTGGATTTGGGGAGAAATGTTTTTACCACACATCCCAAAATCAG 4217

QY 302 ATTTCTCATCAATCCCATGATGATCTATTCACAGAGCATGTCGACGGGTGATGCTCTAG 361
DB 4218 ATTTCTCATCAATCCCATGATGATCTATTCACAGAGCATGTCGACGGGTGATGCTCTAG 4277

QY 362 CAGGCGCTGTGGAGAGAGGCGCCAGCCCAAGTCAAGAGGTGGGTAGGGGCTTCCAGCAC 421
DB 4278 CAGGCGCTGTGGAGAGAGGCGCCAGCCCAAGTCAAGAGGTGGGTAGGGGCTTCCAGCAC 4337

QY 422 AGGCGCTCTCTGTCTGTGGGCAACATGCTCTGCTGTGAGAGCTTTGGCCACGTCCTGCTC 481
DB 4338 AGGCGCTCTCTGTCTGTGGGCAACATGCTCTGCTGTGAGAGCTTTGGCCACGTCCTGCTC 4397

QY 482 ATTTGAGCCTGC 493
DB 4398 ATTTGAGCCTGC 4409

RESULT 2
AK098471 1938 bp mRNA linear PRI 12-SEP-2003
LOCUS AK098471
DEFINITION Homo sapiens cDNA FLJ25605 f18, clone JTH14272.
ACCESSION AK098471
VERSION AK098471.1 GI:21758489
KEYWORDS oligo capping; fig (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.

REFERENCE
AUTHORS 1 Ninomiya, K., Wagatsuma, M., Kanda, K., Kondo, H., Yokoi, T.,
Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K.,
Kamihara, K., Katsuma, N., Sato, K., Tanikawa, M., Yamazaki, M.,
Suzuki, Y., Hata, H., Nakagawa, K., Mizuno, S., Morinaga, M.,
Kawanura, M., Sugiyama, T., Irie, R., Otsubi, T., Sato, H.,
Nishikawa, T., Sugiyama, A., Kawakami, B., Nagai, K., Teagai, T. and
Sugano, S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1938)
AUTHORS Sugano, S. and Suzuki, Y.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-2002) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure, Human Genome
Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail: flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of

Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology (RAB); cDNA library
construction and 5'-end one pass sequencing; Institute of Medical
Science, University of Tokyo, Laboratory of Genome Structure, Human
Genome Center; 3'-end one pass sequencing; RAB; clone selection for
full insert sequencing; RAB and Helix Research Institute.
Location/Qualifiers

FEATURES

source

1. 1938
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="JTH14272"
/cell_line="JCR"
/tissue_type="chryoid"
/clone_lib="JTH"
/note="cloning vector: pME18FLJ3"
237..1706
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC05312.1"
/db_xref="GI:21758490"
/translation="MESSRLKBLTMDLKKDLIGTQERATALLBQVAGSMRD
LLHNSLDRKSKLGLAQHDFGAFEPLOKRLDQVROVAERGLQDLPKQAO
SRLQGLQEBGLDGHQMEARPLVOENPHQKMDQLSDFOALQRSLDLPKQAO
VOEHCCTPSHQLLELROMIVVTTQKLEARGEGEPDASOEAFERLVAEPEKXAO
SLVEAQGLVMEKSPREGAAVVOELRELASGMRRLRLLESLSLIRNMHLORMED
SGKRVPTNNI PKSGFLINPMDPIRHRBRANLLOEBSHSDPSOLIRNPOWLOVE
NSKLVRIIATMTSTREDI.PTRKSKLOELKAPVPEQGH.FENILIRGARGSTDEYEDL
RYOMLVYKSLDKSDGHLTQSSPGEPTQKTRNRGUGSLFRVCCVALPLQLLL
FLILFLPIREDRSCLANNFARSFLIMRNGPPT"

CDS

ORIGIN

Query Match 70.5%; Score 347.6; DB 8; Length 1938;
Best Local Similarity 98.9%; Pred. No. 2e-65;
Matches 350; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGCGGCGCGGGGAGATGCCGAGTCCCAAGAGCGGAGTTTGAGAGGCTGTGGCAGAAATT 61
DB 809 GCGGCGCGCGGGGAGATGCCGAGTCCCAAGAGCGGAGTTTGAGAGGCTGTGGCAGAAATT 868

QY 62 CCGGAGAAAGAGAGCCCAAGCTGTCTCTGTGTGAAGCGCAGGCGCTGGTGTATGAGAA 121
DB 869 CCGGAGAAAGAGAGCCCAAGCTGTCTCTGTGTGAAGCGCAGGCGCTGGTGTATGAGAA 928

QY 122 GTTTTCTCCGAGAGGCTGTCCGTTGTCAGAGAGAGCTGAGGAGCTGGCAGAGTCG 181
DB 929 GTTTTCTCCGAGAGGCTGTCCGTTGTCAGAGAGAGCTGAGGAGCTGGCAGAGTCG 988

QY 182 GCGGCGCTTGAGGCTGTGGAAGAAAGTCTGTGAGCCTCATGAGAACTGGCATCTGCA 241
DB 989 GCGGCGCTTGAGGCTGTGGAAGAAAGTCTGTGAGCCTCATGAGAACTGGCATCTGCA 1048

QY 242 GAGAGTGAAGTGGATTTGGGGAGAAATGTTTTTCAACCAACATCCCAAAGTCAAG 301
DB 1049 GAGAGTGAAGTGGATTTGGGGAGAAATGTTTTTCAACCAACATCCCAAAGTCAAG 1108

QY 302 ATTTCTCATCAATCCCATGATGATCTATTCACAGAGCATGTCGACGGGTGATGCTCTAG 355
DB 1109 ATTTCTCATCAATCCCATGATGATCTATTCACAGAGCATGTCGACGGGTGATGCTCTAG 1162

RESULT 3
CQ735548 3978 bp DNA linear PAT 03-FEB-2004
LOCUS CQ735548
DEFINITION Sequence 21482 from Patent WO02068579.
ACCESSION CQ735548
VERSION CQ735548.1 GI:42327813
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.

REFERENCE 1
 AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
 TITLE K15, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof
 JOURNAL Patent: WO 02068579-A 21482 06-SEP-2002;
 PE Corporation (NY) (US)
 FEATURES
 source location/Qualifiers
 1..3978
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 70.5%; Score 347.6; DB 6; Length 3978;
 Best Local Similarity 98.9%; Pred. No. 1.9e-65;
 Matches 350; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGGCGGCCCCGGGGATGCCAGATCCCAAGGCCGAGTTTGAGAGGCTGTGGCAGAAATT 61
 DB 2847 GGGGGGCCCCGGGGATGCCAGATCCCAAGGCCGAGTTTGAGAGGCTGTGGCAGAAATT 2906
 QY 62 CCGGAGAGAGAGGCCCAAGCTGTCTCTGTGGAGCCGAGGCTGTGGTGGAGAA 121
 DB 2907 CCGGAGAGAGAGGCCCAAGCTGTCTCTGTGGAGCCGAGGCTGTGGTGGAGAA 2966
 QY 122 GTCTTCTCCGAGAGGCTGTCTCTGTGGAGAGCTCAGGAGCTGTGGAGAGTGTG 181
 DB 2967 GTCTTCTCCGAGAGGCTGTCTCTGTGGAGAGCTCAGGAGCTGTGGAGAGTGTG 3026
 QY 182 GCGGCGCTTGAGAGCTGTCTGAGAGAGTCTGAGGCTTCATCGAAATCGCATCTGCA 241
 DB 3027 GCGGCGCTTGAGAGCTGTCTGAGAGAGTCTGAGGCTTCATCGAAATCGCATCTGCA 3086
 QY 242 GAGGATGGAAGTGGATTCGGGGAGAAATGTTTTCACCAACAATCCCAAGTCAAG 301
 DB 3087 GAGGATGGAAGTGGATTCGGGGAGAAATGTTTTCACCAACAATCCCAAGTCAAG 3146
 QY 302 ATTTCTCATCAATCCCATGATCTTATTCAGGAGCATGTGAGCGCTGAGTCT 355
 DB 3147 ATTTCTCATCAATCCCATGATCTTATTCAGGAGCATGTGAGCGCGCAATCT 3200

RESULT 4
 LOCUS CQ845922 3003 bp DNA linear PAT 02-AUG-2004
 DEFINITION Sequence 4569 from Patent EP1440981.
 ACCESSION CQ845922
 VERSION CQ845922.1 GI:50895521
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Isegai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S., Yamamoto, J., Isono, Y., Nagai, K. and Irie, R.
 TITLE Full-length human cdna
 JOURNAL Patent: EP 1440981-A 4569 28-JUL-2004;
 Research Association for Biotechnology (JP)
 FEATURES
 source location/Qualifiers
 1..3003
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 70.1%; Score 345.4; DB 6; Length 3003;
 Best Local Similarity 99.7%; Pred. No. 5.9e-65;
 Matches 346; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGCGGCCCCGGGGATGCCAGATCCCAAGGCCGAGTTTGAGAGGCTGTGGCAGAAATT 61

DB 1911 GGGCGGCCCCGGGGATGCCAGATCCCAAGGCCGAGTTTGAGAGGCTGTGGCAGAAATT 1970
 QY 62 CCGGAGAGAGAGGCCCAAGCTGTCTCTGTGGAGCGCAGGCTGTGGTGGAGAA 121
 DB 1971 CCGGAGAGAGAGGCCCAAGCTGTCTCTGTGGAGCGCAGGCTGTGGTGGAGAA 2030
 QY 122 GTCTTCTCCGAGAGGCTGTCTCTGTGGAGAGCTCAGGAGCTGTGGAGAGTGTG 181
 DB 2031 GTCTTCTCCGAGAGGCTGTCTCTGTGGAGAGCTCAGGAGCTGTGGAGAGTGTG 2090
 QY 182 GCGGCGCTTGAGAGCTGTCTGAGAGAGTCTGAGGCTTCATCGAAATCGCATCTGCA 241
 DB 2091 GCGGCGCTTGAGAGCTGTCTGAGAGAGTCTGAGGCTTCATCGAAATCGCATCTGCA 2150
 QY 242 GAGGATGGAAGTGGATTCGGGGAGAAATGTTTTCACCAACAATCCCAAGTCAAG 301
 DB 2151 GAGGATGGAAGTGGATTCGGGGAGAAATGTTTTCACCAACAATCCCAAGTCAAG 2210
 QY 302 ATTTCTCATCAATCCCATGATCTTATTCAGGAGCATGTGAGCGGG 348
 DB 2211 ATTTCTCATCAATCCCATGATCTTATTCAGGAGCATGTGAGCGGG 2257

RESULT 5
 LOCUS AK131436 3003 bp mRNA linear PRI 07-MAY-2004
 DEFINITION Homo sapiens cDNA FLJ16564 f18, clone SYN04005889.
 ACCESSION AK131436
 VERSION AK131436.1 GI:4707387
 KEYWORDS oligo capping, f18 (full insert sequence).
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Oca, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hiro, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and Isegai, T.
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3003)
 AUTHORS Isegai, T. and Yamamoto, J.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2004) Takao Isegai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamataari, Kisarazu, Chiba 292-0818, Japan (E-mail: genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES
 source location/Qualifiers
 1..3003
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="SYNOV4005889"
 /cissue_type="synovial membrane tissue from rheumatoid arthritis"
 /clone_id="SYNOV4"
 /note="Cloning vector: PME18SFL3"
 79..2793
 /note="unmated protein product"
 /codon_start=1

```
/protein_id="BAB18582.1"
/db_xref="GI:47077388"
/translation="MAEALACCPDQKQKGLIARLKDIQAQNEEYVTTWTHCRSREI
VLMHSEYLLARDEFYRFQKQWVLEPHIELQGLKEQWOMLSHQVILLHVNDAV
LLDRLEBAASLFFNRIGDPSVDEADQKMAEYDAVAKAOKVDLEJEAHSEYQ
AGVDEFLQKAVKAVNGCLGKNGCKLPITORLSTODLAKDPREESELEJEEBSA
GVYRNTSPIGAEEKITGELEEMRKVLKELALAMEERERIRGLRSAGAEKOOIKLEA
ELSEPRMVQRLAEGILOPAKAGTDELVAAHRRYSATRRALASEPRVDLQNALK
ELIVFPHNLKPLSDSVIATTOEYSLKVSARLNAAVAMQHFQRPLODLQMLAL
AORLEVTASLPDLPSLHTFLPQIEALAMESSRLKELTMDLQKOLLIGIQERAT
ALLEVAGSMRBDLLHNSLLOKSKLOSLAQHOKFGAFEPLQKOLLIGIQERAT
KGIQORLPGKQOQLSRLOGLQEBGLDQMEAPARPLVOENPHQKMDLSDFQAL
OSLEBDLVRCQSOVQEHCTFSHOLELRQWIVYMOKLBAHNGEGPDAISOAEF
ELVNAEPPEKEQQLSLVEAQMLVMEKSSPEGAAYOEBELRELAESWRALRIEBSL
SLIRNMHLQRMEDVSGKRVFTWNIKPSGFLINPMDPIFRHRRREESGHEDFSQLR
NFGOMLQVENSKLRIIAMRTSAEDLRKSKLOLEAVRPEQHLFENLRLRGAR
GTSDELEDRLYQWMLYKSLKDSGHLTOSSPPEPFGQKTRMRGLSLFRACCAV
LPLQLLLFLFLFLFLPIREEDRSCTLANNFARSPTMLRVNGPPT"
```

ORIGIN

```
Query Match 70.1%; Score 345.4; DB 8; Length 3003;
Best Local Similarity 99.7%; Pred. No. 5,9e-65;
Matches 346; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGCGGCGGGGGGAGTCCGAGTCCCAAGAGCGGAGTTGAGAGGCTGGGAGAAATT 61
DB 1911 GGGGGGCGGGGGGAGTCCGAGTCCCAAGAGCGGAGTTGAGAGGCTGGGAGAAATT 1970
QY 62 CCGGAGAGAGAGAGAGAGTCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
DB 1971 CCGGAGAGAGAGAGAGAGTCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2030
QY 122 GTCTTCTCCGAGAGAGTCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGTCTG 181
DB 2031 GTCTTCTCCGAGAGAGTCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGTCTG 2090
QY 182 GGGGGGCTTGAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTCTGCA 241
DB 2091 GGGGGGCTTGAGAGTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTCTGCA 2150
QY 242 GAGATGAGAGAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTCA 301
DB 2151 GAGATGAGAGAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTCA 2210
QY 302 ATTTCTCATCATCCCATGAGTCTTATTTCCAGAGCATCGTCAAGCGG 348
DB 2211 ATTTCTCATCATCCCATGAGTCTTATTTCCAGAGCATCGTCAAGCGG 2257

RESULT 6
CNS01DW/c 150350 bp DNA linear PRI 28-NOV-2001
LOCUS Human chromosome 14 DNA sequence BAC C-2588C21 of library Caltech-D
DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL138539
VERSION AL138539.7 GI:117154410
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Molecular Systematics: Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 150350)
Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brocquier, P., Cactolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
Levy, M., Eckenberg, R., Bruls, T., deBardine, V., Cruaud, C.,
Gyapay, G., Saurin, W. and Weissenbach, J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 150350)
Genoscope.
Direct Submission
JOURNAL Submitted (28-NOV-2001) Genoscope - Centre National de Sequencage :
```

COMMENT

BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Nov 29, 2001 this sequence version replaced gi:15796544.
----- Genome Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: segref@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-725G5
Downstream BAC (overlapping the SP6 end) : R-1070N10 (AC=AL133467)

----- Summary Statistics

Assembly program: Phrap, version 2.0

Quality coverage: 14.73x in Q20 bases; sum-of-contigs

Overall quality chart :

```
Range : bases
0 :
1 - 9 :
10 - 19 :
20 - 29 : 15
30 - 39 : 230
40 - 49 : 3272
50 - 59 : 8973
60 - 69 : 8851
70 - 79 : 14405
80 - 89 : 34504
90 - 99 : 80100
```

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES

Location/Qualifiers

1..150350

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="14"

/clone="C-2588C21"

/clone_11b="Caltech-D"

59057..59165

/note="matching EMBL:H91307

RHdb:RH44999

dbSTS:STS38067

Identified using the e-PCR software (G. Schuler)"

59380..59597

/note="matching EMBL:Z67463

RHdb:RH53924

dbSTS:STS42921

Identified using the e-PCR software (G. Schuler)"

107778..107904

/note="matching EMBL:AA447603

RHdb:RH92376

dbSTS:STS65379

Identified using the e-PCR software (G. Schuler)"

ORIGIN

```
Query Match 57.5%; Score 283.4; DB 8; Length 150350;
Best Local Similarity 96.3%; Pred. No. 1.4e-51;
Matches 290; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 193 GGCTGCTGGAAGAAAGTCTGCTGAGCTCATAGAAAGTGGCATGCAAGAGATGGAAG 252
DB 90950 GACTGGCTGTAATTTGTTCTTCACCCCTCATGAAAGTGGCATGCAAGAGATGGAAG 90891
QY 253 TGGATTTCGGGGAAGAAATGATTTTCCACCAACAATCCCAAGTCAGAGATTTCTCATCA 312
DB 90890 TGGATTTCGGGGAAGAAATGATTTTCCACCAACAATCCCAAGTCAGAGATTTCTCATCA 90831
QY 313 ATCCATGATCTTATTTCCAGGATCGTGAAGCGGTGAGTGTCTTACAGAGGCTGTGG 372
DB 90830 ATCCATGATCTTATTTCCAGGATCGTGAAGCGGTGAGTGTCTTACAGAGGCTGTGG 90771
```

QY 373 GAGAGGGGCGAGCCCGCAGGTCAAGAGTGGGTAGGGGTCTCCAGCACAGGCCCTCC 432
 DB 90770 GAGAGGGGCGAGCCCGCAGGTCAAGAGTGGGTAGGGGTCTCCAGCACAGGCCCTCC 90711
 QY 433 TGTCTGGGGGCAACATGCTGTCTGTAGAGCACTTGCCCACTCTCTCTCACTTTGAGCTG 492
 DB 90710 TGTCTGGGGGCAACATGCTGTCTGTAGAGCACTTGCCCACTCTCTCTCACTTTGAGCTG 90651
 QY 493 C 493
 DB 90650 C 90650
 RESULT 7
 LOCUS CNS01DRC 166007 bp DNA linear HTG 10-MAR-2005
 DEFINITION Homo sapiens chromosome 14 clone R-298123, WORKING DRAFT SEQUENCE.
 ACCESSION AL117186
 VERSION AL117186.3 GI:7710964
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.
 1 (bases 1 to 166007)
 Genoscope.
 Direct Submission
 Submitted (23-MAY-2000)
 On May 5, 2000 this sequence version replaced gi:6981774.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in progress
 and the release of this data is based on the understanding that the
 sequence may change as work continue. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phage, etc. . . even if efforts are made to eliminate these
 contaminating sequences. The following BAC sequence is oriented
 from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : R-1070N10
 Downstream BAC (overlapping the SP6 end) : C-2588C21

 Overall quality chart :
 Range : bases
 0 : 4
 1 - 9 : 676
 10 - 19 : 1988
 20 - 29 : 5400
 30 - 39 : 19159
 40 - 49 : 16768
 50 - 59 : 13248
 60 - 69 : 21617
 70 - 79 : 42283
 80 - 89 : 30090
 90 - 99 : 14774

 Percentage of bases with a quality value >= 40 : 83 %.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 166007: contig of 166007 bp in length.
 Location/Qualifiers
 1..166007
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="R-298123"

ORIGIN /clone_1ib="RPCL-11"
 Query Match 57.5%; Score 283.4; DB 14; Length 166007;
 Best Local Similarity 96.3%; Pred. No. 1,4e-51;
 Matches 290; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 QY 193 GGCTGCTGGAAGAAAGTCTGAGCTCATAGAAATGGCATCGCAGAGATGAAG 252
 DB 92267 GACTGGGCTGAGTGTCTCTTCAGCTCATAGAAATGGCATCGCAGAGATGAAG 92326
 QY 253 TGGATTGGGGGAGAAATGTTTTCACCAACAATCCCAAGTACAGATTTTCATCA 312
 DB 92327 TGGATTGGGGGAGAAATGTTTTCACCAACAATCCCAAGTACAGATTTTCATCA 92386
 QY 313 ATCCCATGATCTTATTCCTCCAGGATGTCAGACGCTGATCTGTACAGGGCTGTG 372
 DB 92387 ATCCCATGATCTTATTCCTCCAGGATGTCAGACGCTGATCTGTACAGGGCTGTG 92446
 QY 373 GAGAGGGGCGAGCCCGCAGGTCAAGAGTGGGTAGGGGTCTCCAGCACAGGCCCTCC 432
 DB 92447 GAGAGGGGCGAGCCCGCAGGTCAAGAGTGGGTAGGGGTCTCCAGCACAGGCCCTCC 92506
 QY 433 TGTCTGGGGGCAACATGCTGTCTGTAGAGCACTTGCCCACTCTCTCACTTTGAGCTG 492
 DB 92507 TGTCTGGGGGCAACATGCTGTCTGTAGAGCACTTGCCCACTCTCTCACTTTGAGCTG 92566
 QY 493 C 493
 DB 92567 C 92567
 RESULT 8
 LOCUS BC099694 3370 bp mRNA linear ROD 25-JUL-2005
 DEFINITION Mus musculus cDNA clone MGC:106312 IMAGE:4507746, complete cds.
 ACCESSION BC099694
 VERSION BC099694.1 GI:71043409
 KEYWORDS MGC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 3370)
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
 Altschul,S.F., Collins,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heide,F.,
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
 Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheetz,T.E., Brownstein,M.J., Uebli,T.B., Toshiyuki,S.,
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwen,P.J.,
 McKernan,K.O., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahey,J.J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smalins,D.E.,
 Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 Mammalian Gene Collection Program Team
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16699-16903 (2002)
 JOURNAL PUBMED 12477932
 2 (bases 1 to 3370)
 NIH MGC Project
 Direct Submission
 Submitted (15-JUL-2005) National Institutes of Health, Mammalian

REMARK
COMMENT

Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@hgrl.nih.gov
Akhter, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, O.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stancijop, S., Thomas, P.J., Touchman, J.W.,
Taugen, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRAC Plate: 208 Row: d Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 27359681.

FEATURES
source

1. 3370
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="MGC:106312 IMAGE:4507746"
/tissue_type="eye, retina, mouse strain C57BL/6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
293. 3220
/codon_start=1
/product=Unknown (protein for MGC:106312)"
/protein_id="AAH9694.1"
/db_xref="GI:71043410"
/translation="MTQGPDEPERSVEDAQAAMKVIYIQOLQVNDTKPRALAEAL
RETERKICOLESGMVKVLELRAAALLATCOGKPELIRALGDKKQMEVETWT
HNSRTEVWMLHSEVILLNODERFQKQVAVLREPVLOGLDKKQKSHAOVLT
HNDVNOVLDRLLEAGSLFSGIPSVDEQKQKABYDVKARQRDLDLQV
ADHEQYRDVNEFQMLKAVKHSCLGRNCKLATELSTLOIADPFGESL
KRLEQAVGIVONTSPLEAKISGELEKGVLEKLVLMKEEGRLGLQSGDCE
OOIOLEAEIGDFKSLQRLAOGLEPTVATKEDLVNQMLFSGTRALASEPRV
DRIOQLKLVPEPDLOSLSVATIOEQSMKGNTRLHNTAELMORORPLND
LOIKMLAQRLLDITSLPDLASITFLPQIEALRESSRLKQMLADLKTDLGSI
FQGERATLLEQVTSVRDLDLHNSLQKSLQSLVLOHNDFGAFLPRLKLDL
QATIOAEKGLPDLPGKQVQLRLQGLQEGDLQVITAVRPLAGNSKQKQVQI
SCQOALQSLLELDVRCQNVNEHCTFSHRLSELOLMTAOTLESQGVRLMDA
ESQEALETLLSEIPEKEVQSLQALQGLVMKSPREGATVQOEELRLKMSWQAR
LLEENMLSLMRNQLORTQVDTGKQVFTNNIPKAGFLINPOPIPRQHGNSPEGH
DUPEDPQLLRDFEQMLQANSLKRLIITMRVATADLPTREYKLOELARPEQHL
FEHLRLPRADPSNELEDRTKMLFKSLKDSGHLTSSGELTARQKSRQKRM
SPSLQKACRYALPLQILLFLLLFLPLPAGEERSALANNFARSPALMLRYNGP
PPT"

ORIGIN

Query Match 43.0%; Score 212.2; DB 9; Length 3370;
Best Local Similarity 77.8%; Pred. No. 6.1e-36;
Matches 256; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 7 GCCCGGGGATGCGAGTCCCAAGAGCCGAGTTTGAAGAGCTGTGGCAGATTCCTCCG 66
DB 2325 GTCTGTGGATGCTGAGTCCCAAGAGCTGAGTCCGAGAGCTGCTGTCTGGAATCCAG 2384
QY 67 AGAAGAGGCCAGCTGTCTCCCTGTGGAGACGAGGCTGGCTGTGTGATGAGAGACTTT 126
DB 2385 AGAAAGAGGTCAGAGTGTCTCTCCAGACATGGGCGAGCTTGTGATGAGAGACTTT 2444

QY 127 CTCGGAGGATGTCGCCGTGGTGCAGAGAGACCTCAGGAGCTGCGACAGTCTGTCGCG 186
DB 2445 CCCCAAGAGGGCAACCATGTGTCAGAGAGACTTGAAGTGTGATCTTTGGCAGG 2504
QY 187 CCTTGAAGCTGTGTCGAGAGAAAGTCTGTGAGCTTCATGAGAAATCTGATCTGCAGAGA 246
DB 2505 CCTCGGCTGTGTAAGAGGAAACATGCTGATGATCATGAGAAACAGCAGCTGCAGAGA 2564
QY 247 TGGAGTGTGATTCGGGGAAGAAATGCTTTTACCAACAACATCCCAAGTCAGATTTT 306
DB 2565 CAGAGGTGACACGGGGAAGAGGAGTTCACCAACAACATCCCAAGCGGCTTTT 2624
QY 307 TCATCATCCATGATGATTCATTCCTATTCACG 335
DB 2625 TCATCAACCTTGAAGACCCCATTCACG 2653

RESULT 9
AC152545/c
LOCUS
DEFINITION
AC152545 165217 bp DNA linear HTG 01-JUL-2005
Bos taurus clone CH240-3M23, WORKING DRAFT SEQUENCE, 16 unordered
pieces.
AC152545
AC152545.4 GI:68227057
VERSION
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Bos taurus (cow)
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 165217)
REFERENCE
AUTHORS
Muzny, D., Marie, Metker, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D.,
Aryalpech, V., Ayogaji, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaramaie, D., Barber, M., Barnstead, M., Barnham, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Cesaar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Cleveland, M., Davis, C., Davy-Carroll, J., De And, C., Dederich, D.,
Devlado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, T., Guevara, W.,
Gunnarane, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M.,
Hollins, B., Howell, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpaty, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kows, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Louisedge, H., Lozdo, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmood, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundaasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, G., Olariunagoun, A., Pal, S.,
Nwackeme, O., Okunnu, G., Olariunagoun, A., Pal, S., Parke, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primm, E., Pu, L., Reigh, R.,
Puzo, M., Quito, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, M., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sison, I., Sittler, C.D., Smaj, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K.,

Valae, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Meinkoth, G., and Gibbs, R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 165217)
Worley, K.C.
Direct Submission
Submitted (12-NOV-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 165217)
Cow Genome Sequencing Consortium.
Direct Submission
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jun 26, 2005 this sequence version replaced g1:58037954.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Project name: PAJW
Center name: CH240-3M23
Center Clone name: Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 155449 bases at least Q40
Consensus quality: 156707 bases at least Q30
Consensus quality: 157568 bases at least Q20
Estimated insert size: 157550; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length.
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a working draft sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 33332: contig of 33332 bp in length
* 3333 33382: gap of 50 bp
* 33383 46869: contig of 13487 bp in length
* 46870 46919: gap of 50 bp
* 46920 54137: contig of 7218 bp in length
* 54138 54695: gap of 558 bp
* 54696 72807: contig of 18112 bp in length
* 72808 74316: gap of 1509 bp
* 74317 76686: contig of 2370 bp in length
* 76687 76765: gap of 79 bp
* 76766 89950: contig of 13185 bp in length
* 89951 91560: gap of 1610 bp
* 91561 110253: contig of 18693 bp in length

FEATURES
Source
1..165217
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone="CH240-3M23"
33333..33382
/estimated_length=50
46870..46919
/estimated_length=50
54138..54695
/estimated_length=558
72808..74316
/estimated_length=1509
76687..76765
/estimated_length=79
89951..91560
/estimated_length=1610
110254..110303
/estimated_length=50
120119..120554
/estimated_length=436
131920..131969
/estimated_length=50
136598..136697
/estimated_length=unknown
138983..139032
/estimated_length=50
147526..148107
/estimated_length=582
149196..149295
/estimated_length=unknown
161576..163037
/estimated_length=1462
164067..164166
/estimated_length=unknown

ORIGIN
Query Match 28.5%; Score 140.6; DB 14; Length 165217;
Best Local Similarity 72.7%; Pred. No. 1.8e-20;
Matches 210; Conservative 0; Mismatches 74; Indels 5; Gaps 2;

QY 208 GTCGCTGAGCGCTTATCGAAATCGCATCTGCAGAGATG---GAAGTGGATTCCGGGGA 264
DB 138160 GACTTCTCAGCTCACCAGAAACCGGACGTCGACAGATGACGAAAGTGGATTCAAGGA 138101
QY 265 AGAAATGTTTTCACCAACAATCCCAAGTCAGGATTTTCATCAATCCATGATC 324
DB 138100 AGAAATGATTTCACCAACAATCCCAAGTGGGTTTCTCATCTCCCGACGGATC 138041
QY 325 CTATTCACGAGCATGTCGACCGCTGAGTCTGTTTACAGGCGCTGTGGAGAAAGGGCCA 384
DB 138040 TTACTCCCGGATCGCGCTGAGTGAAGTCAATCGGCTTTCAGGAGGGCCGAGA 137981
QY 385 GGCCCCAGTCAAGAGTGGGTAGGGGTCTCCACACACAGGCCCTCCCTGCTGAGGACAA 444

Dn	137980	TCCGAGCTCAGTTGATGGTGGGAAGGTGTACCGACGAAAGCCCCCTTCCTCCCATCTGGGGCCA	1379221
Oy	445	CATGCTCTGCCTCTGAGGACTTGSGCACGTCCTGTCTCATTTGAGCTGC	493
Dn	137920	CATGCTCAGCTC--AGGACTGGGGCCCTGTCTGTCTTGGGTGAGTGTGC	137874
RESULT 10			
ACI47375/c			
LOCUS	ACI47375	176510 bp	DNA linear ROD 29-MAY-2004
DEFINITION	Mus musculus BAC clone RP24-159013 from chromosome 12, complete sequence.		
VERSION	ACI47375		
KEYWORDS	ACI47375.3 GI:47564188		
SOURCE	HTG.		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.		
AUTHORS	Levy, A. and Kozlowski, A.		
JOURNAL	The Sequence of Mus musculus BAC clone RP24-159013 Unpublished (2001)		
REFERENCE	2 (bases 1 to 176510)		
AUTHORS	Wilson, R.K.		
JOURNAL	Submitted (11-NOV-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	3 (bases 1 to 176510)		
AUTHORS	Wilson, R.K.		
JOURNAL	Submitted (17-DEC-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	4 (bases 1 to 176510)		
AUTHORS	Wilson, R.K.		
JOURNAL	Submitted (22-MAY-2004) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	5 (bases 1 to 176510)		
AUTHORS	Wilson, R.K.		
JOURNAL	Submitted (29-MAY-2004) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
COMMENT	On May 22, 2004 this sequence version replaced gi:39975582.		
	----- Genome Center		
	Center: Washington University Genome Sequencing Center		
	Center code: WUGSC		
	Web site: http://genome.wustl.edu		
	Contact: submissions@watson.wustl.edu		
	----- Summary Statistics		
	Center project name: M_BB0159013		

NOTICE:			
	This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., paired quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.		
	MAPPING INFORMATION: Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu		
	SOURCE INFORMATION:		

The RPCL-22BAC library has been constructed by Plietser de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Plietser de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is overlapped by AC133077 and AC140477.

FEATURES

Source

repeat_region	/organelle="Mus musculina /mol_type="genomic DNA" /db_xref="taxon:10090" /chromosome="12" /map="12" /clone="RP24-159013" /clone1bb="RPci-24" 813. ,964 1144. ,1387 /rptc_family="B4" 1394. ,1716 /rptc_family="MaLR" 1722. ,2860 /rptc_family="MaLR" 2861. ,3004 /rptc_family="Alu" 3005. ,3117 /rptc_family="MaLR" 3123. ,3536 /rptc_family="MaLR" 4400. ,4646 /rptc_family="B4" 5441. ,5554 5674. ,5732 /rptc_family="MER1_type" 5733. ,5918 /rptc_family="B2" 5919. ,5937 /rptc_family="RMBR1B" 6186. ,6331 /rptc_family="Alu" 6461. ,6507 /rptc_family="MIR" 7024. ,7144 /rptc_family="MIR" 7619. ,7686 /rptc_family="B4" 7870. ,8087 /rptc_family="B4" 9190. ,9254 /rptc_family="MIR" 10468. ,10828 /rptc_family="L1" 11586. ,11664 /rptc_family="ERVX" 12182. ,12555 /rptc_family="L1" 13446. ,13458 /rptc_family="MaLR" 13459. ,13537 /rptc_family="Alu" 13538. ,13562 /rptc_family="MaLR" 13678. ,13682 /rptc_family="B2" 13894. ,13938 /rptc_family="MER1_type" 14211. ,14367 /rptc_family="B4" 14431. ,14650 /rptc_family="B2"
---------------	--

```
repeat_region      14666. .14853
                    /rpc_family="B2"
repeat_region      15844. .16144
                    /rpc_family="B4"
tRNA               15847. .15917
                    /product="tRNA-Val"
repeat_region      16202. .16286
                    /rpc_family="MIR"
repeat_region      18142. .18324
                    /rpc_family="B2"
repeat_region      18744. .18997
                    /rpc_family="B4"
repeat_region      19955. .20107
                    /rpc_family="B4"
repeat_region      20367. .20381
                    /rpc_family="Alu"
repeat_region      20382. .20530
                    /rpc_family="Alu"
repeat_region      20531. .20641
                    /rpc_family="Alu"
repeat_region      21096. .21381
                    /rpc_family="B4"
repeat_region      22347. .22445
                    /rpc_family="L1"
repeat_region      22472. .22533
                    /rpc_family="ERV1"
repeat_region      24262. .24977
                    /rpc_family="L1"
repeat_region      25408. .25593
                    /rpc_family="B2"
repeat_region      25605. .25680
                    /rpc_family="B4"
repeat_region      25718. .25792
                    /rpc_family="ID"
repeat_region      26024. .26158
                    /rpc_family="MER1_type"
repeat_region      27153. .27297
                    /rpc_family="MER1_type"
repeat_region      28246. .28543
                    /rpc_family="MER2_type"
repeat_region      29941. .30039
                    /rpc_family="Alu"
repeat_region      30808. .31101
                    /rpc_family="B4"
repeat_region      31418. .31467
                    /rpc_family="ERV1"
repeat_region      31652. .31777
                    /rpc_family="MALR"
repeat_region      32253. .32353
                    /rpc_family="ACHobo"
repeat_region      33136. .33381
                    /rpc_family="MIR"
repeat_region      33617. .33662
                    /rpc_family="Alu"
repeat_region      33711. .33763
                    /rpc_family="ID"
repeat_region      34334. .34710
                    /rpc_family="MALR"
repeat_region      36098. .36315
                    /rpc_family="B4"
repeat_region      36650. .36735
                    /rpc_family="MIR"
repeat_region      39916. .40251
                    /rpc_family="RMER1B"
Query Match      24.6%; Score 121.4; DB 9; Length 176510;
Best Local Similarity 69.8%; Pred. No. 2.8e-16;
Matches 178; Conservative 0; Mismatches 76; Indels 1; Gaps 1;
```

```
QY      208 GCTCTGCTGAGCTCTCATCAGAAATCGCATCTGCAGCATGGAATGCAATTGCGGGAGA 267
DB      160158 GCTCTCCCACTCATGAGAAACGACGCTGCAGACAGAGGTGACACGCGGAGAGA 160099
```

```
QY      268 AATGTGTTTCCACCAACATCCCAAGTCAGATTTTCATCATCAATCCATGATCTCA 327
DB      160098 AGCAGGTTCACCAACATCCCAAGCCGCTTTCATCAACCTTCAGAGACCCA 160039
QY      328 TTCCAGCATCGTCGACCGGTGATGTTCTTACGACGGCTGTGGAGAGAGGCCAGGC 387
DB      160038 TTCCAGGACACGACATGGGGTATGTCGTCTCCAGGTTTGGAGA-ACGTCAGATCC 159980
QY      388 CCCAGGTCAAGAGTGGTGGGTTCCAGCAGAGCCCTTCCTGTCTGGGGCAACAT 447
DB      159979 AATGTCAAGTAGAGAGGTGGGGCTCCAGCACAGTTCCTTCGCTGATCGCAT 159920
QY      448 GCTCTGCTGAGCA 462
DB      159919 GCTCAGCTCATGGCA 159905

RESULT 11
AP008207_101/c
WPCOMMENT
Sequence split into 433 fragments LOCUS AP008207 Accession AP008207
Fragment Name      Begin      End
AP008207_000      1      11000
AP008207_001      100001      210000
AP008207_002      200001      310000
AP008207_003      300001      410000
AP008207_004      400001      510000
AP008207_005      500001      610000
AP008207_006      600001      710000
AP008207_007      700001      810000
AP008207_008      800001      910000
AP008207_009      900001      1010000
AP008207_010      1000001      1110000
AP008207_011      1100001      1210000
AP008207_012      1200001      1310000
AP008207_013      1300001      1410000
AP008207_014      1400001      1510000
AP008207_015      1500001      1610000
AP008207_016      1600001      1710000
AP008207_017      1700001      1810000
AP008207_018      1800001      1910000
AP008207_019      1900001      2010000
AP008207_020      2000001      2110000
AP008207_021      2100001      2210000
AP008207_022      2200001      2310000
AP008207_023      2300001      2410000
AP008207_024      2400001      2510000
AP008207_025      2500001      2610000
AP008207_026      2600001      2710000
AP008207_027      2700001      2810000
AP008207_028      2800001      2910000
AP008207_029      2900001      3010000
AP008207_030      3000001      3110000
AP008207_031      3100001      3210000
AP008207_032      3200001      3310000
AP008207_033      3300001      3410000
AP008207_034      3400001      3510000
AP008207_035      3500001      3610000
AP008207_036      3600001      3710000
AP008207_037      3700001      3810000
AP008207_038      3800001      3910000
AP008207_039      3900001      4010000
AP008207_040      4000001      4110000
AP008207_041      4100001      4210000
AP008207_042      4200001      4310000
AP008207_043      4300001      4410000
AP008207_044      4400001      4510000
AP008207_045      4500001      4610000
AP008207_046      4600001      4710000
AP008207_047      4700001      4810000
AP008207_048      4800001      4910000
AP008207_049      4900001      5010000
AP008207_050      5000001      5110000
```

AP008207_051 5100001 5210000
AP008207_052 5200001 5310000
AP008207_053 5300001 5410000
AP008207_054 5400001 5510000
AP008207_055 5500001 5610000
AP008207_056 5600001 5710000
AP008207_057 5700001 5810000
AP008207_058 5800001 5910000
AP008207_059 5900001 6010000
AP008207_060 6000001 6110000
AP008207_061 6100001 6210000
AP008207_062 6200001 6310000
AP008207_063 6300001 6410000
AP008207_064 6400001 6510000
AP008207_065 6500001 6610000
AP008207_066 6600001 6710000
AP008207_067 6700001 6810000
AP008207_068 6800001 6910000
AP008207_069 6900001 7010000
AP008207_070 7000001 7110000
AP008207_071 7100001 7210000
AP008207_072 7200001 7310000
AP008207_073 7300001 7410000
AP008207_074 7400001 7510000
AP008207_075 7500001 7610000
AP008207_076 7600001 7710000
AP008207_077 7700001 7810000
AP008207_078 7800001 7910000
AP008207_079 7900001 8010000
AP008207_080 8000001 8110000
AP008207_081 8100001 8210000
AP008207_082 8200001 8310000
AP008207_083 8300001 8410000
AP008207_084 8400001 8510000
AP008207_085 8500001 8610000
AP008207_086 8600001 8710000
AP008207_087 8700001 8810000
AP008207_088 8800001 8910000
AP008207_089 8900001 9010000
AP008207_090 9000001 9110000
AP008207_091 9100001 9210000
AP008207_092 9200001 9310000
AP008207_093 9300001 9410000
AP008207_094 9400001 9510000
AP008207_095 9500001 9610000
AP008207_096 9600001 9710000
AP008207_097 9700001 9810000
AP008207_098 9800001 9910000
AP008207_099 9900001 10010000
AP008207_100 10000001 10110000
AP008207_101 10100001 10210000
AP008207_102 10200001 10310000
AP008207_103 10300001 10410000
AP008207_104 10400001 10510000
AP008207_105 10500001 10610000
AP008207_106 10600001 10710000
AP008207_107 10700001 10810000
AP008207_108 10800001 10910000
AP008207_109 10900001 11010000
AP008207_110 11000001 11110000
AP008207_111 11100001 11210000
AP008207_112 11200001 11310000
AP008207_113 11300001 11410000
AP008207_114 11400001 11510000
AP008207_115 11500001 11610000
AP008207_116 11600001 11710000
AP008207_117 11700001 11810000
AP008207_118 11800001 11910000
AP008207_119 11900001 12010000
AP008207_120 12000001 12110000
AP008207_121 12100001 12210000
AP008207_122 12200001 12310000
AP008207_123 12300001 12410000

AP008207_124 12400001 12510000
AP008207_125 12500001 12610000
AP008207_126 12600001 12710000
AP008207_127 12700001 12810000
AP008207_128 12800001 12910000
AP008207_129 12900001 13010000
AP008207_130 13000001 13110000
AP008207_131 13100001 13210000
AP008207_132 13200001 13310000
AP008207_133 13300001 13410000
AP008207_134 13400001 13510000
AP008207_135 13500001 13610000
AP008207_136 13600001 13710000
AP008207_137 13700001 13810000
AP008207_138 13800001 13910000
AP008207_139 13900001 14010000
AP008207_140 14000001 14110000
AP008207_141 14100001 14210000
AP008207_142 14200001 14310000
AP008207_143 14300001 14410000
AP008207_144 14400001 14510000
AP008207_145 14500001 14610000
AP008207_146 14600001 14710000
AP008207_147 14700001 14810000
AP008207_148 14800001 14910000
AP008207_149 14900001 15010000
AP008207_150 15000001 15110000
AP008207_151 15100001 15210000
AP008207_152 15200001 15310000
AP008207_153 15300001 15410000
AP008207_154 15400001 15510000
AP008207_155 15500001 15610000
AP008207_156 15600001 15710000
AP008207_157 15700001 15810000
AP008207_158 15800001 15910000
AP008207_159 15900001 16010000
AP008207_160 16000001 16110000
AP008207_161 16100001 16210000
AP008207_162 16200001 16310000
AP008207_163 16300001 16410000
AP008207_164 16400001 16510000
AP008207_165 16500001 16610000
AP008207_166 16600001 16710000
AP008207_167 16700001 16810000
AP008207_168 16800001 16910000
AP008207_169 16900001 17010000
AP008207_170 17000001 17110000
AP008207_171 17100001 17210000
AP008207_172 17200001 17310000
AP008207_173 17300001 17410000
AP008207_174 17400001 17510000
AP008207_175 17500001 17610000
AP008207_176 17600001 17710000
AP008207_177 17700001 17810000
AP008207_178 17800001 17910000
AP008207_179 17900001 18010000
AP008207_180 18000001 18110000
AP008207_181 18100001 18210000
AP008207_182 18200001 18310000
AP008207_183 18300001 18410000
AP008207_184 18400001 18510000
AP008207_185 18500001 18610000
AP008207_186 18600001 18710000
AP008207_187 18700001 18810000
AP008207_188 18800001 18910000
AP008207_189 18900001 19010000
AP008207_190 19000001 19110000
AP008207_191 19100001 19210000
AP008207_192 19200001 19310000
AP008207_193 19300001 19410000
AP008207_194 19400001 19510000
AP008207_195 19500001 19610000
AP008207_196 19600001 19710000

AIOQMPSKRSYIEVQALAIQWISLISFELSKAERSLIVASVALAIVETGEPKQPVYF
 SLAQMGPSSRYIEIEKLAVLAWLAKRLGHYQKHKVIVIPSOYPIGEILIRDKQVYF
 SKRAALSTPPIHFVRAIPAKISQVTLADPAVETSVLAPEPPEAEOSSWIMCSGWSHK
 GAGIAVLESFNGVPIPIRYARLQDPDTTNAAEYATLTLRKAQALGRRLLIRIDSK
 LVAGHAIEKSPFAKEEGMKRYLLEAVRSMEECFGIYEHLPROMNEADVALASTVGA
 PHSQPIFFVYVLAHPMSQDSSEYMAIDQEKLESDPDWMTPEVKHLETGMILVDEAA
 KILQIATQIKWYSGMOLYRSQVLOPLLRISFVEGKEMAKIHQGLCGHQAIVYAS
 KYRQGVWMPYLVKVCVEQIKCESCQGRSGAQAQVLOLQIAPITPWPAPGPIIG
 PVPVANGKPAIVNVEYSPRTAEPLCAITSAAYQWVIMNYICRPREVPEPIITG
 GKQFSDKREKEGELNLEIRFASVALDSNGAERTNGKITIHALKRLLEGATKQKP
 BELLISVLWVRPPOQPSLMTAPALGANSPPWMSGGEEERVSLELLEGEVREAL
 EHLHKATSTSATYNNKVRPTBMLPGLVLRKKANPVAAGVESKWEQPYLLIKHNSRT
 GSFHALTEEGEEDHSMAASLKRFPVY"
 complement (19612. . 20131)
 /note="5' UTR"
 gene
 LTR
 complement (join (22358. . 23037, 23572. . 23841, 24555. . 24948))
 /gene="OSJNBA0004B13.5"
 complement (join (22358. . 23037, 23572. . 23841, 24555. . 24948))
 /gene="OSJNBA0004B13.5"
 /note="contains ESTs AU032146 (R3628), AU065568 (C62863)"
 /product="putative tubulin beta-4 chain"
 /codon_start=1
 /protein_id="BAB39951.1"
 /db_xref="gi:13486716"
 /translation="MREILHIQGQGGCQNIQSGKFWVCGDEHGIDPTGKVTSLQTL
 ERYVNYNNAASGGRPYRPAVLMDLEBGTMDSRVTGQITRPDPNFVQGSAGNNMA
 KKHVYEGALILSDLVVRKEANECDCLOGFQVCHSLGGGTSGTSGMVLITSRIRFEP
 DDMMLFVSFSPKSYSDIVPEYRNATLSHVQLEVNADCEVINDNALYDI CRTKLKT
 TSPGQIHLNISATWSGVCCLLRPOLNSDLKLAIVNIPPPRLFFVWAGAPLTSR
 GSOQVYLVPELTQOMDAKNMCAADPRHGRYVLASMPFGKSTKEVDEOMINVO
 NNSQSYFVNIPINNYKSSVCDIPRGLASAPRTFIGNSTSGMFPKRVSGOPFAMVRK
 AELHWYTGCGMEMETTESMNNDLVSYEQYQDRTADEBDYDEDEEQVEDE"
 join (29477. . 29557, 30999. . 31115, 31504. . 31614)
 /gene="OSJNBA0004B13.6"
 join (29477. . 29557, 30999. . 31115, 31504. . 31614)
 /gene="OSJNBA0004B13.6"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB39952.1"
 /db_xref="gi:13486717"
 /translation="MDCODLVQKIIIGVRASTLLASNGSIIIGHVEGRPOKODLIVSIF
 TSPAVQULLIHADSXGLSHRLSHNNILAMNGTAVRGAMRSRTIDRTVTFLLPLRDV
 "
 complement (join (37256. . 38344, 38502. . 38665, 41064. . 41640))
 /gene="OSJNBA0004B13.7"
 complement (join (37256. . 38344, 38502. . 38665, 41064. . 41640))
 /gene="OSJNBA0004B13.7"
 /note="hypothetical protein
 similar to Arabidopsis thaliana chromosome 5, KMC9.22"
 /codon_start=1
 /protein_id="BAB39953.1"
 /db_xref="gi:13486718"
 /translation="MASPFRSLPLPILLVFPVSLFVYLLLHHRSSIPAEELSGPG
 PBPBRSLILKYLAVDRPGLRCLRLSLAAAYADRVLAHLVADHPEDNSLDSRE
 ILAEADALMPHGEKRYHTRANAGIQAWITAMNPGSDERLAVPDDLEKSPILYR
 FLKRLVMAIYYDRENSPYVFGASLQRPFFVAGKGNKTIQDSETHLFLYOVNWGQ
 LILPKXWKEALRYDEHSEKSGIKPILEGKTIQWYKXGERTWMTPEIIFKFNHSCYFN
 FTFNFKKEBALSVSHHDAGVNGRSVGPSTTLDDKNDLPNLEIMELQIPLNKIKWYQF
 AVLNPBRVIRKFSSELGYSVOLENNVLITLSLGEORIANLITCYLEKSGMRVYIF
 LYNDPBRLLDLAHRGHPYDAISLOSITMSSSIYSDPVKEITVYKAVIYIKRDLG
 NLYLVANGTISLGSKLNEPDSQVDFPAESYDLNLFRLKRSQSSKRTNMBLDIRADG
 NMSSSGFSSLEHKNFVNLVGLVGNNAQVGLKDEITAVALEGPTTSNKSISEGH
 CMLFMSHMTSDVOSQLENNGLWLIISDSSCSAVVCGQKK"
 join (42066. . 42180, 43982. . 44125, 45133. . 45258, 45484. . 45602
 45939. . 46208)
 /gene="OSJNBA0004B13.8"
 join (42066. . 42180, 43982. . 44125, 45133. . 45258, 45484. . 45602
 45939. . 46208)
 /gene="OSJNBA0004B13.8"
 /note="contains ESTs AU062927 (C51629), AU030693 (E60120)
 similar to Arabidopsis thaliana chromosome 1, F14P1.8
 unknown protein"
 /codon_start=1

Query Match 11.6%; Score 57; DB 15; Length 142268;
Best Local Similarity 55.2%; Pred. No. 0.03;
Matches 111; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Query 1 CGGCGGCGCCCGGGGGAGATGCCGAGTCCCAAGAGCCGAGTTTGAGAGGCTGGTGCAGAAAT 60
Db 104620 CGACGCGAGCTTGGCGAGAACCTTGTCCTTGGAGAGGTCGAGCGGACATGCTGAGGCGGAG 1045
61 TCCCGGAGAGAGAGGCCCGACCTGCTCCTGGTTCGAGAGCGGAGGCTGGCTGTGATGATGAGA 120
Db 104560 AGCCGAGAGGATGTGTGCACAGAGGTGGCGAGAGCTGAGAGAAATGTGTGAGAGG 1045
Qy 121 AGCTTTTCGAGAGGCTGTCGCGTGGTGCAGAGAGGAGCTCAGGAGCTGCAGAGTCGT 180
Db 104500 ACGTCCCGAGGAAACCGAGAGATGTGTGAGAGAGATGCGGAGCCGAGAGAGATGC 1044
Qy 181 GCGCGGCGCTTGAAGCTGCTGG 201
Db 104440 TGACACGAGATGCCGCTGTG 104420

RESULT 13
AP003631/c 154248 bp DNA linear PLN 22-DEC-2004
LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
DEFINITION PAC clone:PO581P09.
ACCESSION AP003631 BA000010
VERSION AP003631.3 GI:15408812
KEYWORDS Oryza sativa (japonica cultivar-group)
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Burkholderia, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
Spermatophyta, Magnoliophyta, Liliopsida, Poales, Poaceae,
Ehrhartoideae, Oryzaceae, Oryza.
1
Sasaki,T., Matsumoto,T., Yamamoto,K., Sakata,K., Baba,T.,
Katayose,Y., Wu,J., Nishimura,Y., Cheng,Z., Nagamura,Y.,
Antonino,B.A., Kanamori,H., Hosokawa,S., Nasukawa,M., Arikawa,K.,
Chiden,Y., Hayashi,M., Okamoto,M., Ando,T., Aoki,H., Arita,K.,
Hamada,M., Harada,C., Hijishita,S., Honda,M., Ichikawa,Y.,
Idonuma,A., Iijima,M., Ikeda,M., Ikeno,M., Ito,S., Ito,T., Ito,Y.,
Ito,Y., Imwubuchi,A., Kamiya,K., Karasawa,W., Katagiri,S.,
Kikuta,A., Kobayashi,N., Kono,I., Machita,K., Meehara,T.,
Mizuno,H., Mizubayashi,T., Muka,Y., Nagasaki,H., Nakashima,M.,
Nakama,Y., Nakemichi,Y., Nakamura,M., Namiki,N., Negishi,M.,
Ohta,I., Ono,N., Saij,S., Sakai,K., Shibata,M., Shimokawa,T.,
Shomura,A., Song,J., Takazaki,Y., Terasawa,K., Teuji,K., Waki,K.,
Yamagata,H., Yamane,H., Yoshiki,S., Yoshizawa,R., Yokawa,K.,
Zhong,H., Iwama,H., Endo,T., Ito,H., Hahn,J.H., Kim,H.I., Eun,M.Y.,
Yano,M., Jiang,Y. and Gojobori,T.
The genome sequence and structure of rice chromosome 1
Nature 420 (6913), 312-316 (2002)
12447438
2 (base 1 to 154248)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Subassembly
Submitted (17-MAY-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki, 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Aug 31, 2001 this sequence version replaced gi:14861131.
Genes predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
(http://www.softberry.com/), Genemark.hmm
(http://opal.biology.gatech.edu/genemark/), Glimmer
(http://cigr.org/cdb/glimmerm/glimr_form.html), RiceHMM

COMMENT
JOURNAL
AUTHORS
PUBMED
TITLE
REFERENCE
JOURNAL

CDS
 join(24883..24998,25151..25535)
 /gene="P0581F09.8"
 /note="Predicted by GeneMark.hmm etc."
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="BAD82605.1"
 /db_xref="GI:56785023"
 /translation="MATATYATDSELYVTVYPKDAIPDDDDGGGAASRGRMLRPSLLV
 GDLPLPADGEGRRGRDAAEGERRGMDASCGERRGRDSSAGEERRRMDAGVGE
 SASEDEPHRDVMSGAPARMTCSRRTYRSVPAAASPVHPSHAMPREXERRRGREREGEK

Query Match 11.6%; Score 57; DB 15; Length 154248;
 Best Local Similarity 55.2%; Pred. No. 0.03;
 Matches 111; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 1 CGGCGCGCCCGGGGATGCCGAGTCCCAAGAGCCGAGTTTGAGAGCGTGTGGCAAT 60
 DB 127490 CGACGAGCTGGCGAGACCTTGTGGGAGGTGCAGCGCGACATGTGTGAGCGGAGG 127431
 QY 61 TCCCGGAGAAAGAGGCCCACTCTCCCTGTGTGAAGCGCGAGGCTGGCTGTATGAGA 120
 DB 127430 AGCGGAGAGATGTGTGACAGACAGAGTGGCGAGAGCTGGAGAGATGTGTGACAGAGG 127371
 QY 121 AGCTTCTCCGAGAGGTGCTGCCGTGTGAGAGAGAGCTCAGAGAGCTTGACAGATCGT 180
 DB 127370 ACCTGCGGAGAGAACCGAGAGATGTGTGAGAGAGATGGCGGAGCCGAGAGATGG 127311
 QY 181 GCGCGGCTTTGAGCGTGTCTGG 201
 DB 127310 TGCACGAGATGCCGCTTGTGG 127290

RESULT 14
 AP003020/c
 LOCUS
 DEFINITION
 Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
 PAC clone: P0498A12.
 ACCESSION
 AP003020 BA000010
 VERSION
 AP003020.2 GI:13486738
 KEYWORDS
 Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriatoideae; Oryzaceae; Oryza.

REFERENCE
 AUTHORS
 1 Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T.,
 Katayose, Y., Wu, J., Nishimura, Y., Cheng, Z., Nagamura, Y.,
 Antonio, B. A., Kanemori, H., Hosokawa, S., Masukawa, M., Arikawa, K.,
 Chiden, Y., Hayashi, M., Okamoto, M., Ando, T., Aoki, H., Arita, K.,
 Hamada, M., Harada, C., Hijiishi, S., Honda, M., Ichikawa, Y.,
 Idonuma, A., Iijima, M., Ikeda, M., Ikemoto, M., Ito, S., Ito, T., Ito, Y.,
 Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, M., Katagiri, S.,
 Kikuta, A., Kobayashi, N., Kono, I., Machida, K., Maehara, T.,
 Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nagashima, M.,
 Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M.,
 Ohta, I., Ono, N., Saji, S., Sakai, K., Shibata, M., Shimokawa, T.,
 Shomura, A., Song, J., Takazaki, Y., Terasawa, K., Teuji, K., Waki, K.,
 Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K.,
 Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J. H., Kim, H. I., Eun, M. Y.,
 Yano, M., Jiang, J. and Gojobori, T.
 The genome sequence and structure of rice chromosome 1
 Nature 420 (6913), 312-316 (2002)

TITLE
 JOURNAL
 PUBMED
 12447438
 2 (bases 1 to 159749)
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Direct Submission
 Submitted (06-DEC-2000) Takuji Sasaki, National Institute of
 Agricultural Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki, 305-8602, Japan
 (E-mail:tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 On Mar 28, 2001 this sequence version replaced gi:11602827.

COMMENT

Genes were predicted from the integrated results of the following:
 GENSCAN (<http://CCR-081.mit.edu/GENSCAN.html>), EGENESH
 (<http://www.softberry.com/>), GeneMark.hmm
 (<http://opal.biology.gatech.edu/GeneMark/>), GlimmerM
 (http://www.tigr.org/tcb/glimmerm/glmr_form.html), RiceHMM
 (<http://rgp.dna.affrc.go.jp/RiceHMM/>), SplicePredictor
 (<http://bioinformatics.laestate.edu/cgi-bin/8p.cgi>), sam4
 (<http://globin.cse.psu.edu/html/doc/s4ind.html>), gap2
 (<http://www.tigr.org/software/glimmer/>), BLASTN and BLASTX. The
 genomic sequence was searched against NCBI Nonredundant Protein
 database, nr (<ftp://ncbi.nlm.nih.gov/blast/db>) and the cDNA
 sequence database at RGP or DBJ. Protein homologs of the coding
 regions were searched against NCBI Nonredundant Protein database
 with BLASTP. ESTs represent the identified cDNA sequences using
 BLASTN with the corresponding DBJ accession no. and RGP clone ID.
 Full-length cDNAs represent the identified cDNA sequences using
 BLASTN with the corresponding DBJ accession no.
 A gene with identity or significant homology to a protein is
 classified based on the protein name to indicate the homology level
 such as same name, 'putative-' and '-like protein'. A gene without
 significant homology to any protein but with full-length cDNA or
 EST homology (covering almost the entire length of partial
 sequence) is classified as an 'unknown' protein. A gene predicted
 by two or more gene prediction programs is classified as a
 'hypothetical' protein according to IRGSP standard. A gene
 predicted by a single gene prediction program is also classified as a
 probable 'hypothetical' protein and is included as a
 miscellaneous feature of the sequence.
 The orientation of the sequence is from SP6 to T7 of the PAC clone.
 This sequence of P0498A12 clone has an overlap with P0581F09 (DBJ:
 AP003631) clone at 5' end and with P0511C01 (DBJ: AP002070) clone
 at 3' end. Detailed information on overlap and assembly quality
 together with annotation of this entry is available at
<http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

FEATURES

source
 1..159749
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="1"
 /clone="P0498A12"
 1..333
 /gene="P0498A12.1"
 1..>333
 /note="start and end point are not identified"
 1..333
 /gene="P0498A12.1"
 /note="contains full-length cDNA(s): AK119993"
 /codon_start=1
 /product="unknown protein"
 /protein_id="BAD81552.1"
 /db_xref="GI:56784167"
 /translation="MEMKSYSSSEKPKEDTISRATSPSSSSSAGTRPSCWRS
 VTTTARRRCCGDEHDTAARASVLTGHDHLLGICGVDSAGAVRTLVERNDPYC
 DECNTHGR"
 complement(join(1116..2215,3205..3549))
 /gene="P0498A12.2-1"
 complement(join(1116..2215,3205..3549))
 /gene="P0498A12.2-1"
 /note="supported by full-length cDNA(s): AK064158"
 complement(join(1424..2215,3205..3465))
 /gene="P0498A12.2-1"
 /note="contains EST(s): AU166951 (CG0192), AU166950 (CG0192)
 contains full-length cDNA(s): AK064158
 similar to Arabidopsis thaliana chromosome 3, AC3g17615"
 /codon_start=1
 /product="unknown protein"
 /protein_id="BAB39974.1"
 /db_xref="GI:13486740"
 /translation="MGAGMSSGRRSPFYGGGGGCVGRPGMLPLALQVLLLEYGRAG
 ASRPVTAALLAANLTYLRPGSLDALLPSINRVAFNPHLLIHYCDLTFPLSAYHL

/note="predicted by GeneMark.hmm etc."
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="BAD81554.1"
 /db_xref="GI:56784169"
 /translation="MGAVGDSSESDRVRNGENHEARSSLGDGGPRKVGGEREAT
 ALVPGRTGSGVGNKATTTTKTGAFGLWSVNASERIVILIGYSSLSKG"
 23593.
 /gene="P0498A12.8"
 /note="P0498A12.8"
 join(<22938..23027,23399..23486,23586..23593)
 /gene="P0498A12.8"
 /note="start and end point are not identified"
 join(22938..23027,23399..23486,23586..23593)
 /gene="P0498A12.8"
 /note="predicted by GENSCAN etc."

Query Match 11.6%; Score 57; DB 15; Length 159749;
 Best Local Similarity 55.2%; Pred. No. 0.03;
 Matches 111; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Oy 1 CGCGCGGCGCGCGGATGCCGATGCCCAAGAGCGCCGACTTTGAAGAGCTGTGTGCCAGAT 60
 Db 45335 CGACGAGACTTGGCGGACCTTGTCTTGGAGGTGCGCGGACATGTGTGAGCGGAGG 45227
 Oy 61 TCCGGGAGAGAGAGCGCCAGCTGTCCCGTGGTGGAGAGCGCAGGCTGTGTGTGATGAGAG 120
 Db 45275 AGCGGAGAGATGTGTGAGCAGAGAGTGTGCGGAGAGCTGAGAGATGTGTGACAGAG 45227
 Oy 121 AGCTTCTCCGAGAGGTGCTCCGTGTGTGACAGAGAGCTGAGAGCTGACAGATGCT 180
 Db 45215 ACGTGCGCGAGAGACCGAGAGAGATGTGTGCGAGAGATGTGCGGAGCGGAGAGATG 45155
 Oy 181 GCGGCGCTTGAGGCTGCTG 201
 Db 45155 TGACGAGATGCCGCTGTTG 45135

RESULT 15 2279 bp mRNA linear INV 10-JUN-200
 LOCUS AY571308
 DEFINITION Kukulcania hibernalis major ampullate spideroin 1 mRNA, partial cds
 ACCESSION AY571308
 VERSION AY571308.1 GI:47007922
 KEYWORDS
 SOURCE
 ORGANISM
 Kukulcania hibernalis
 Kukulcania hibernalis
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 Araneomorphae; Haplogynae; Filistatidae; Kukulcania.
 1 (bases 1 to 2279)
 Tian,M., Liu,C. and Lewis,R.
 Analysis of major ampullate silk cDNAs from two non-orb-weaving
 spiders
 JOURNAL Biomacromolecules 5 (3), 657-660 (2004)
 PUBMED 15132643
 REFERENCE 2 (bases 1 to 2279)
 AUTHORS Liu,C., Tian,M. and Lewis,R.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAR-2004) Molecular Biology, University of Wyoming,
 1000 E. University Avenue, Laramie, WY 82071-3944, USA
 FEATURES
 source
 1..2279
 /organism="Kukulcania hibernalis"
 /mol_type="mRNA"
 /db_xref="taxon:268415"
 <1..>2279
 /codon_start=1
 /product="major ampullate spideroin 1"
 /protein_id="AA084433.1"
 /db_xref="GI:47007923"
 /translation="CGSSGDSAAASAAAAAGSGRGGPGAGARSGAGAGAGAGS
 VCGYSGGAGAGAGAGAGGEGFGBGCGYGAAGAGYAGRGAGRGAGRGAG
 SASASAVVFESAGAGAGYGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
 GRGAGGGAFAFSASASAVVFESAGPGEAGSGSGAGASAAAAAGAGSGRGP

CDS
 115

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 03:01:41 ; Search time 466 Seconds
(without alignments)
7050.844 Million cell updates/sec

Title: US-10-071-510A-16
Perfect score: 493
Sequence: 1 cggcgcggccgggagatgcc.....cctgtctcatttgagcctgc 493

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues
Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
1: Genesegq1980s:*
2: Genesegq1990s:*
3: Genesegq2000s:*
4: Genesegq2001as:*
5: Genesegq2001bs:*
6: Genesegq2002as:*
7: Genesegq2002bs:*
8: Genesegq2003as:*
9: Genesegq2003bs:*
10: Genesegq2003ds:*
11: Genesegq2004as:*
12: Genesegq2004bs:*
13: Genesegq2005as:*
14: Genesegq2005bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	493	100.0	493	10	ADH69132 Human tum
2	347.6	70.1	4083	5	AAS83954 DNA encod
3	345.4	70.1	3003	12	ADG67408 Novel hum
4	176.4	35.8	532	12	ACH77474 Human gen
5	174	35.3	174	12	ACH91174 Human gen
6	56.8	11.5	2282	8	AAU51693 Kukulcani
7	51.6	10.5	3711	13	ADU01598 Novel hum
8	51.4	10.4	1505	2	AAO55750 Genomic c
9	50.2	10.2	15231	6	ABK63598 Rat seque
10	50.2	10.2	15231	10	ADBS59205 Toxicity-
11	50.2	10.2	15231	10	ADBS59206 Primary r
12	50.2	10.2	15231	10	ADT42078 Toxicity
13	50.2	10.2	15231	10	ADP72968 Renal tox
14	49.8	10.1	768	6	ABZ13007 Arabidops
15	49.8	10.1	2328	6	ADU01973 Novel hum
16	49.8	10.1	16442	2	AAK72613 Human imm
17	49.4	10.0	8298	4	AAK72613 Human imm
18	49	9.9	8973	12	ADP28653 Human sec
19	48.6	9.9	114793	4	AAO08215 Human gen

20	48.2	9.8	628	12	ACH89664 Human gen
21	47.8	9.7	799	2	AAV55831 Nucleotid
22	47.8	9.7	913	2	AAU07189 Cotton fl
23	47.8	9.7	913	2	AAU13034 Cotton fl
24	47.8	9.7	913	2	AAU130253 Cotton fl
25	47.8	9.7	913	2	AAU130253 Cotton fl
26	47.8	9.7	913	2	AAU130253 Cotton fl
27	47.8	9.7	913	2	AAU130253 Cotton fl
28	47.8	9.7	913	2	AAU130253 Cotton fl
29	47.8	9.7	913	2	AAU130253 Cotton fl
30	47.8	9.7	913	2	AAU130253 Cotton fl
31	47.8	9.7	913	2	AAU130253 Cotton fl
32	47.8	9.7	913	2	AAU130253 Cotton fl
33	47.8	9.7	913	2	AAU130253 Cotton fl
34	47.8	9.7	913	2	AAU130253 Cotton fl
35	47.8	9.7	913	2	AAU130253 Cotton fl
36	47.8	9.7	913	2	AAU130253 Cotton fl
37	47.8	9.7	913	2	AAU130253 Cotton fl
38	47.8	9.7	913	2	AAU130253 Cotton fl
39	47.8	9.7	913	2	AAU130253 Cotton fl
40	47.8	9.7	913	2	AAU130253 Cotton fl
41	47.8	9.7	913	2	AAU130253 Cotton fl
42	47.8	9.7	913	2	AAU130253 Cotton fl
43	47.8	9.7	913	2	AAU130253 Cotton fl
44	47.8	9.7	913	2	AAU130253 Cotton fl
45	47.8	9.7	913	2	AAU130253 Cotton fl

ALIGNMENTS

RESULT 1
ADH69132 standard; DNA; 493 BP.
ID ADH69132;
AC ADH69132;
AC 25-MAR-2004 (first entry)
DT Human tumour resistance/sensitivity marker DNA #16.
DE human; ds; tumour; tumour growth; cancer; resistance; sensitivity.
KW Homo sapiens.
XX US2003143552-A1.
XX 31-JUL-2003.
XX 08-FEB-2002; 2002US-00071510.
XX 08-FEB-2001; 2001US-0267276P.
XX (MILL-) MILLENNIUM PHARM INC.
XX Clark E, Grenfell-Lee T, Lu K, Hartmann L, Brown JL;
XX WPI; 2003-829783/77.
XX Determining whether agent can/cannot be used to reduce growth of tumor
XX involves exposing obtained tumor cells to test agent, determining
XX expression level of markers by tumor cells.
XX Disclosure; SEQ ID NO 16; 41pp; English.
XX The invention relates to a method of determining whether an agent
XX can/cannot be used to reduce the growth of tumour. involves obtaining a
XX sample of tumour cells, exposing them to one or more test agents,
XX determining the level of expression of one or more markers by the cells
XX exposed to the agent and by the cells not exposed to the agent and
XX identifying the agent as appropriate/inappropriate to reduce the growth
XX of tumour. The method is useful for determining whether an agent
XX can/cannot be used to reduce the growth of tumour. The method is useful

CC for determining whether an anti-cancer agent should in treatment of a
CC cancer patient should be continued/discontinued. The method is useful for
CC reducing growth rate of cancer in a patient. The present sequence
CC represents a human tumour resistance/sensitivity DNA marker.
XX

Sequence 493 BP, 99 A; 124 C; 174 G; 96 T; 0 U; 0 Other;

Query Match 100.0%; Score 493; DB 10; Length 493;
Best Local Similarity 100.0%; Pred. No. 7, 2e-114;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CGGCGCGCGCGCGGATCCGAGTCCCAAGAGCCGAGTTTGAGAGCTGTGGCAGAAT 60
DB 1 CGGCGCGCGCGCGGATCCGAGTCCCAAGAGCCGAGTTTGAGAGCTGTGGCAGAAT 60
QY 61 TCCCGGAGAGAGAGCCGAGCTGTCCCTGTGAGAGCGCAGGCTGTGTATGAGAGA 120
DB 61 TCCCGGAGAGAGAGCCGAGCTGTCCCTGTGAGAGCGCAGGCTGTGTATGAGAGA 120
QY 121 AGCTTCTCCGAGGAGGCTGCGTGTGTGAGAGAGAGCTCAGGAGCTGGCAGAGCTGT 180
DB 121 AGCTTCTCCGAGGAGGCTGCGTGTGTGAGAGAGAGCTCAGGAGCTGGCAGAGCTGT 180
QY 181 GCGCGGCTTGAAGGCTGTGAGAGAGAGAGCTGTGAGCTCATCAGAACTGGCATCTGC 240
DB 181 GCGCGGCTTGAAGGCTGTGAGAGAGAGAGCTGTGAGCTCATCAGAACTGGCATCTGC 240
QY 241 AGAGGATGAGAGTGTGATTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 241 AGAGGATGAGAGTGTGATTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 301 GATTTCATCAATCCATGATGATCTTATCCAGGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 301 GATTTCATCAATCCATGATGATCTTATCCAGGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 361 GCAGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 361 GCAGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 421 CAGGCGCTTCCCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 421 CAGGCGCTTCCCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 CATTGAGAGCTGC 493
DB 481 CATTGAGAGCTGC 493
```

RESULT 2

AAS83954
AAS83954 standard; cDNA; 4083 BP.

AC AAS83954;

DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #19758.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001MO-US008631.

PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX
PI Drmanac RT, Liu C, Tang YT;

XX WPI, 2001-639362/73.

DR P-PSDB; ABG19767.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PS Claim 1; SEQ ID NO 19758; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 4083 BP, 923 A; 1177 C; 1228 G; 755 T; 0 U; 0 Other;

Query Match 70.5%; Score 347.6; DB 5; Length 4083;

Best Local Similarity 98.9%; Pred. No. 4, 2e-77;

Matches 350; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 2 GCGCGGCGCGGAGAGTGCAGATCCCAAGAGCCGAGTTTGAAGAGCTGTGGCAGAATT 61
DB 2796 GCGCGGCGCGGAGAGTGCAGATCCCAAGAGCCGAGTTTGAAGAGCTGTGGCAGAATT 2855
QY 62 CCGGAGAGAGAGGCGCAGCTGTCCCTGTGTGAGAGCGCAGGCTGTGTGATGAGAGA 121
DB 2856 CCGGAGAGAGAGGCGCAGCTGTCCCTGTGTGAGAGCGCAGGCTGTGTGATGAGAGA 2915
QY 122 GTCCTTCCGAGAGGTCGTCGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 181
DB 2916 GTCCTTCCGAGAGGTCGTCGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2975
QY 182 GCGGCGCTTGAAGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
DB 2976 GCGGCGCTTGAAGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3035
QY 242 GAGATGAGAGTGTGATTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
DB 3036 GAGATGAGAGTGTGATTCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3095
QY 302 ATTTCATCAATCCATGATGATCTTATCCAGGAGAGAGAGAGAGAGAGAGAGAGAG 355
DB 3096 ATTTCATCAATCCATGATGATCTTATCCAGGAGAGAGAGAGAGAGAGAGAGAGAG 3149
```

RESULT 3

AD067408
AD067408 standard; cDNA; 3003 BP.

AC AD067408;

DT 07-OCT-2004 (first entry)

XX Novel human cDNA sequence #2381.
 DE
 XX
 XX ss: gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
 KW cytoabatic; gene therapy; diagnostic marker; morbid state; osteoporosis;
 KW neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
 KW cancer.
 XX
 XX Homo sapiens.
 OS
 XX EP1440981-A2.
 PN
 XX 28-JUL-2004.
 PD
 XX 21-JAN-2004; 2004EP-00001196.
 PF
 XX 21-JAN-2003; 2003JP-00102206.
 PR 09-MAY-2003; 2003JP-00131392.
 XX
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.
 PA
 XX Ieogai T, Sugiyama T, Oseuki T, Makamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Nagai K, Irie R,
 XX WPI: 2004-535376/52.
 DR P-PSDB; ADQ67715.
 XX
 XX Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
 PT
 XX
 XX Claim 1; SEQ ID NO 4569; 2449pp; English.
 PS
 XX The invention relates to 2495 novel polynucleotides (I) and their encoded
 CC polypeptides, sequences hybridizing to these nucleotides, sequences
 CC encoding partial polypeptides and sequences having 70% or 90% identity to
 CC the nucleotide and protein sequences. The nucleotides and polypeptides
 CC are useful as diagnostic markers or therapeutic target for the diseases
 CC or morbid states. They are also useful for treating osteoporosis,
 CC neurological diseases, Alzheimer's diseases, Parkinson's diseases,
 CC dementia and various cancers. This sequence corresponds to a nucleotide
 CC sequence of the invention.
 CC
 XX
 XX Sequence 3003 BP; 664 A; 817 C; 988 G; 534 T; 0 U; 0 Other;
 SQ
 Query Match 70.1%; Score 345.4; DB 12; Length 3003;
 Best Local Similarity 99.7%; Pred. No. 1.4e-76;
 Matches 346; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 GCGCGGCGCGGGGATGCGCGAGTCCCAAGAGCGCGAGTTTGAAGAGCTGTGTGCAAGATT 61
 DB 1911 GCGCGGCGCGGGGATGCGCGAGTCCCAAGAGCGCGAGTTTGAAGAGCTGTGTGCAAGATT 1970
 QY 62 CCGCGAGAGAGAGCGCCAGCTGTCTCTGTGTGAGCGCAAGGCTGCTGTGTGTGAGAA 121
 DB 1971 CCGCGAGAGAGAGCGCCAGCTGTCTCTGTGTGAGCGCAAGGCTGCTGTGTGTGAGAA 2030
 QY 122 GTCTTCTCCGAGGGTGTCTCCGTGTGTGAGAGAGCTCAAGGAGCTGGCAGATCGTG 181
 DB 2031 GTCTTCTCCGAGGGTGTCTCCGTGTGTGAGAGAGCTCAAGGAGCTGGCAGATCGTG 2090
 QY 182 GCGGCGCTTGAAGGCTGTGTGAGAGAGAGTCTGTGAGCTTCATCAGAAATCGGCATCTGCA 241
 DB 2091 GCGGCGCTTGAAGGCTGTGTGAGAGAGAGTCTGTGAGCTTCATCAGAAATCGGCATCTGCA 2150
 QY 242 GAGATGGAAGTGTATTCGGGGAGAAAGATGTTTTCACCAACATCTCCCAATTCAGG 301
 DB 2151 GAGATGGAAGTGTATTCGGGGAGAAAGATGTTTTCACCAACATCTCCCAATTCAGG 2210
 QY 302 ATTTCATCATCCATCCATGATCTATTCCAGGAGCATGTCGACGCG 348
 DB 2211 ATTTCATCATCCATCCATGATCTATTCCAGGAGCATGTCGACGCG 2257

RESULT 4
 ACH77474/C
 ID ACH77474 standard; DNA; 532 BP.
 XX
 AC ACH77474;
 XX
 DT 29-JUL-2004 (first entry)
 XX
 DE Human genome derived single exon probe #10665.
 XX
 KW Human; probe; ss; gene expression; single exon probe; microarray;
 KW alternative splicing event; genomic alteration.
 XX
 OS Homo sapiens.
 OS
 XX US2003194704-A1.
 PN
 XX 16-OCT-2003.
 PD
 XX 03-APR-2002; 2002US-00029386.
 PF
 XX 03-APR-2002; 2002US-00029386.
 PR
 XX (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.
 XX
 PI Penn SG, Rank DR, Hanzel DK;
 DR WPI: 2004-119264/12.
 XX
 XX New human genome-derived single exon nucleic acid probes useful for human
 PT gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 XX
 XX Claim 15; SEQ ID NO 10669; 80pp; English.
 PS
 XX The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridizes under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising at least 8
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid
 CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of selling and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subscription, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterizing
 CC alternative splicing events, in detecting and characterizing gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in printing the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docID=20030194704

XX Sequence 532 BP; 113 A; 175 C; 147 G; 97 T; 0 U; 0 Other;
 SQ Query Match 35.8%; Score 176.4; DB 12; Length 532;
 Best Local Similarity 99.4%; Pred. No. 2,4e-34;
 Matches 177; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 40 TTGAGAGCTGGTGGCAAGTTCCTCCGAGAGAGGCCCAAGCTGCTCCCTGTGGAAGCCG 99
 DB 425 TTGAGAGCTGGTGGCAAGTTCCTCCGAGAGAGGCCCAAGCTGCTCCCTGTGGAAGCCG 366
 QY 100 AGGCGTGGCTGTGATGAGAGAAGTCTTCTCCGAGAGGCTGCTGCTGTGTCAGAGAGAGC 159
 DB 365 AGGCGTGGCTGTGATGAGAGAAGTCTTCTCCGAGAGGCTGCTGCTGTGTCAGAGAGAGC 306
 QY 160 TCAGAGAGCTGGCAGAGTCTGTGGCGGCGCTTGAAGCTGCTGGAAGAAAGTCTGCTAG 217
 DB 305 TCAGAGAGCTGGCAGAGTCTGTGGCGGCGCTTGAAGCTGCTGGAAGAAAGTCTGCTAG 248
 RESULT 5
 ACH91174/c
 ID ACH91174 standard; DNA; 174 BP.
 AC ACH91174;
 XX 29-JUL-2004 (first entry)
 DT Human genome derived single exon probe #24369.
 DE Human genome derived single exon probe; microarray;
 XX Human; probe: ss; gene expression; single exon probe; microarray;
 KM alternative splicing event; genomic alteration.
 XX Homo sapiens.
 OS
 PN US2003194704-A1.
 XX 16-OCT-2003.
 PD 03-APR-2002; 2002US-00029386.
 XX 03-APR-2002; 2002US-00029386.
 PR 03-APR-2002; 2002US-00029386.
 XX (PENN/) PENN S G.
 PA (RANK/) RANK D R.
 PA (HANZ/) HANZEL D K.
 XX Penn SG, Rank DR, Hanzel DK;
 PI WPI; 2004-119264/12.
 DR New human genome-derived single exon nucleic acid probes useful for human
 PT gene expression analysis, for identifying or characterizing alternative
 PT splicing events, for assessing genomic alterations or as tools for
 PT surveying tissues.
 XX Claim 1; SEQ ID NO 24369; 80pp; English.
 PS The invention relates to a nucleic acid probe for measuring human gene
 CC expression, comprising any of the 27,400 fully defined nucleotide
 CC sequences in the specification, or their complements or fragments, and
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences
 CC fully defined in the specification. The probe is a single exon probe that
 CC hybridizes under high stringency conditions to a nucleic acid molecule
 CC expressed in human cells or tissues. Also included are a spatially-
 CC addressable set of single exon nucleic acid probes for measuring human
 CC gene expression (comprising a plurality of single exon nucleic acid
 CC probes cited above, where each of the plurality of probes is separately
 CC and addressably isolatable or amplifiable from the plurality), a single
 CC exon microarray for measuring human gene expression, a method of
 CC measuring human gene expression, a vector comprising the single exon
 CC probe cited above, an ORF-encoded peptide comprising at least 8
 CC contiguous amino acids of any of the above-mentioned amino acid

CC sequences (optionally with conservative amino acid substitutions), an
 CC isolated antibody that binds specifically to a peptide cited above,
 CC methods of gelting and/or licensing single exon probes or microarrays to
 CC a customer desiring to measure gene expression, a method of providing
 CC human gene expression data by subscrption, and a computer-readable
 CC storage medium which contains a database having a plurality of records
 CC (each record including data on the expression of a single exon probe
 CC cited above. The probe, methods and apparatus are useful in gene
 CC expression analysis. The probes may be used as tools for surveying
 CC tissues to detect the presence of expressed messages that contain their
 CC specific exon, or in constructing genome-derived single exon microarrays.
 CC In addition, the probes are used in identifying and characterizing
 CC alternative splicing events, in detecting and characterizing gross
 CC alterations in the genomic locus that includes their exon, in assessing
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,
 CC or in expressing the ORF-encoded peptide. The present sequence is a human
 CC single exon probe of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html?docID=20030194704
 CC
 XX Sequence 174 BP; 31 A; 76 C; 36 G; 31 T; 0 U; 0 Other;
 SQ Query Match 35.3%; Score 174; DB 12; Length 174;
 Best Local Similarity 100.0%; Pred. No. 7.2e-34;
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 44 GAGCGTGGTGGCAAGTTCCTCCGAGAGAGGCCCAAGCTGCTCCCTGTGGAAGCCAGAG 103
 DB 174 GAGCGTGGTGGCAAGTTCCTCCGAGAGAGGCCCAAGCTGCTCCCTGTGGAAGCCAGAG 115
 QY 104 CTGGCTGTGATGAGAGAAGTCTTCTCCGAGAGGCTGCTGCTGTGTCAGAGAGACTCAG 163
 DB 114 CTGGCTGTGATGAGAGAAGTCTTCTCCGAGAGGCTGCTGCTGTGTCAGAGAGACTCAG 55
 QY 164 GGAAGCTGGCAGAGTCTGTGGCGGCGCTTGAAGCTGCTGGAAGAAAGTCTGCTAG 217
 DB 54 GGAAGCTGGCAGAGTCTGTGGCGGCGCTTGAAGCTGCTGGAAGAAAGTCTGCTAG 1
 RESULT 6
 AAL51693
 ID AAL51693 standard; cDNA; 2282 BP.
 XX AAL51693;
 AC AAL51693;
 XX 01-MAY-2003 (first entry)
 DT Kukulcania hibernalis spider silk protein coding sequence #1.
 XX Kukulcania hibernalis spider silk protein; fabric; suture;
 DE Gene; ss; spider silk; spider silk protein; high-tech clothing; rope; reinforced plastic.
 KM medical covering; high-tech clothing; rope; reinforced plastic.
 XX Kukulcania hibernalis.
 OS WO200299082-A2.
 PN 12-DEC-2002.
 PD 06-JUN-2002; 2002WO-US018256.
 XX 06-JUN-2001; 2001US-0296184P.
 PR (UYW-) UNITV WYOMING.
 XX Roth DA, Lewis RV;
 PI WPI; 2003-140616/13.
 XX P-PSDB; AAO16494.
 DR Expressing spider silk protein in a higher plant, by contacting a plant
 PT cell with silk protein encoding a gene linked to a gene that confers
 PT resistance to selection agent, and selecting cells that survive when

Db 1726 GAGAGAGAGAGGTGATGTCGAGAGAGAGTGTCTATGATGTCGAGAGAGAGCTG 1785
|||
Qy 210 CTGCTG 215
|||
Db 1786 GTGATG 1791
|||

RESULT 8
AAQ55750
ID AAQ55750 standard; DNA; 1505 BP.
XX
AC AAQ55750;
XX
DT 25-MAR-2003 (revised)
DT 23-JUL-1994 (first entry)
XX
DE Genomic clone G11F, includes a Brassica root specific promoter.
XX
KM Plants; transformation; food crops; toxins; insects; fungal; infection;
KM ss.
XX
OS Brassica napus.
XX
FH Key Location/Qualifiers
FT CAAAT_signal /tag= .440
FT /tag= b
FT TATA_signal 485..488
FT /tag= c
FT misc_feature 552..554
FT /tag= a
FT /note= "translation start site"
XX
FT
XX
PN WO9402619-A1.
XX
PD 03-FEB-1994.
XX
PF 12-JUL-1993; 93WO-US006541.
XX
PR 16-JUL-1992; 92US-00915246.
XX
PA (PION-) PIONEER HI-BRED INT INC.
XX
PI Beszczynski CJ, Fallis L, Bellemare G, Bolvin R;
XX
DR WPI; 1994-048877/06.
XX
PT Brassica root-specific promoter sequence and vectors containing it - for
PT expression of e.g. toxins in the roots of crops susceptible to root
PT diseases.
XX
PS Disclosure; Fig 4; 54pp; English.
XX
XX The sequence shows the G11F genomic clone isolated from DNA from the
CC roots of Brassica napus and containing a promoter sequence upstream from
CC the translation start site. The promoter enhances gene expression in
CC roots and can be used to confer disease resistance or immunity upon a
CC susceptible plant root by transforming the plant with a composite gene
CC capable of abundant expression in the gene, comprising the promoter and
CC the gene of interest. This method can be used to transform plants, partic
CC food crops that have roots susceptible to fungal or insect diseases. The
CC roots of the transformed plants express a toxin or substance that
CC provides immunity or resistance to the fungal or insect diseases. See
CC also AAQ55749 and AAQ56913-5. (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
SQ Sequence 1505 BP; 395 A; 201 C; 528 G; 381 T; 0 U; 0 Other;
XX
Query Match 10.4%; Score 51.4; DB 2; Length 1505;
Best Local Similarity 49.8%; Pred. No. 0.0079;
Matches 130; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
Qy 14 GGATGCCGAGTCCCAAGAGCGCCGAGTTTGAAGAGCTGTGGCAGAAATTCGCCGAGAAAGA 73

Db 852 GGTATTAGAGAGTGGAAAGCGGTGAAGTGTGTGAGCTGGATTACGAGAGCGCAGAGCTGGA 911
|||
Qy 74 GGCCCAAGCTGTCCTCGTGTGGAAGCCGAGGCTGCTGCTGATGAGAACTCTTTCGGA 133
|||
Db 912 GCACATGTCGTGAGAGAGCGAGAGCGTGTGTGTGAGAGCTGGCCGTCCGAT 971
|||
Qy 134 GGGTCCTGCCGTGTGTCAGAGAGAGCTCAAGAGGCTGCGAGTGTGGCGGCTTGAG 193
|||
Db 972 GGTGTGTGATACGTGTGTGAGAGAGTGTGTGTGTGTGTGAGAGATATGAGAGTGGCGGT 1031
|||
Qy 194 GCTGCTGAAGAAGATCTGCTGAGCCTCATCAGAACTGGCATTCGACAGAGATGAAGT 253
|||
Db 1032 GCAGGTGACATGAGAGTGTGTGAGAGCGGTGGAATGAGAGCGGTGAGAGAGTGAAGT 1091
|||
Qy 254 GGATTCGGGGAGAAAGAAATGCT 274
|||
Db 1092 GCACACGTCGTGTGATACCGT 1112
|||

RESULT 9
ABK63598
ID ABK63598 standard; cDNA; 15231 BP.
XX
AC ABK63598;
XX
DT 18-JUN-2002 (first entry)
XX
DE Rat sequence differentially expressed in response to a hepatotoxin #1505.
XX
KM Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
KM differential expression; centrilobular necrosis; steatosis.
XX
OS Rattus norvegicus.
XX
PN WO200210453-A2.
XX
PD 07-FEB-2002.
XX
PF 30-JUL-2001; 2001WO-US023872.
XX
PR 31-JUL-2000; 2000US-0222040P.
PR 02-NOV-2000; 2000US-0244880P.
PR 11-MAY-2001; 2001US-0290029P.
PR 15-MAY-2001; 2001US-0290645P.
PR 22-MAY-2001; 2001US-0292336P.
PR 06-JUN-2001; 2001US-0295798P.
PR 13-JUN-2001; 2001US-0297457P.
PR 19-JUN-2001; 2001US-0298884P.
PR 09-JUL-2001; 2001US-0303459P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter MW, Johnson KR, Caetle AL, Elashoff MR;
XX
DR WPI; 2002-241625/29.
XX
PT Predicting toxic effects of compounds or the progression of these toxic
PT effects by determining the changes in gene expression in tissues or cells
PT exposed to the toxin and comparing these to gene expression in unexposed
PT tissues or cells.
XX
PS Claim 1; SEQ ID NO 1505; 239pp; English.
XX
XX The invention relates to methods for predicting toxic effects of
CC compounds or the progression of these toxic effects by determining the
CC global changes in gene expression in tissues or cells exposed to the
CC toxin and comparing these to gene expression in unexposed tissues or
CC cells. Also included are methods of predicting at least one toxic effect
CC of a compound or progression of a toxic effect, preferably the
CC hepatotoxicity of a compound, comprising detecting the level of
CC expression in a tissue or cell sample exposed to the compound of two or
CC more genes listed in the specification, where differential expression of

the genes is indicative of at least one toxic effect or progression. The method can also be used to identify an agent which modulates the toxic response and predict cellular pathways that a compound modulates in a cell. The methods utilize a set of at least two probes (on a solid support in kit form), where each of the probes comprises a sequence that specifically hybridizes to a gene listed in the specification, a computer system comprising a database containing information identifying the expression level in a tissue or cell sample exposed to a hepatotoxin of a set of genes comprising at least two genes listed in the specification, and a user interface to view the information used to present information identifying the expression level in a tissue or cell of at least one gene listed in the specification. The method is useful for elucidating global changes in gene expression and for identifying toxicity markers in tissues or cell exposed to a known toxin. The genes may be used as toxicity markers in drug screening and toxicity assays. The genes and gene expression information may be used as diagnostic markers for the prediction or identification of the physiological state of tissue or cell sample that has been exposed to a compound or agent. Hepatotoxicity is characterized by centrilobular necrosis and steatosis. The present sequence is an expressed sequence tag (EST) or cDNA derived from a gene which is differentially expressed in response to a hepatotoxic agent

Sequence 15231 BP; 3451 A; 4277 C; 4932 G; 2571 T; 0 U; 0 Other:

Query Match 10.2%; Score 50.2; DB 6; Length 15231;
Best Local Similarity 49.8%; Pred. No. 0.029;
Matches 127; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

Qy 20 CGAGTCCCAAGAGCCGAGTTTGAAGGCTGGTGCAGAAATCCCGAGAAAGAGGCCCA 79
Db 7947 CGAGCAGAGAGAGAGCCGAGTTTGAAGGCTGGTGCAGAAATCCCGAGAAAGAGGCCCA 8006

Qy 80 GCTGTCCCTGTGAAGCGGAGGCTGGTGCAGAAATCCCGAGAGGCTGC 139
Db 8007 GCTGTCCCTGTGAAGCGGAGGCTGGTGCAGAAATCCCGAGAGGCTGC 8066

Qy 140 TGCCGTGTGAGAGAGGCTGCAGAGCTGGTGCAGAGGCTGGTGCAGAGGCTGC 199
Db 8067 TGCCGTGTGAGAGAGGCTGCAGAGCTGGTGCAGAGGCTGGTGCAGAGGCTGC 8126

Qy 200 GGAAGAAAGTCTGTGAGCTTCATCAGAACTGGCATTCGAGAGATGAGATGATTC 259
Db 8127 GCAAGAGAACTGCAAGCTTCGAGCAGCAGCGGAGAGGAGAACTACTGAGAGA 8186

Qy 260 GGGGAGAAATGCT 274
Db 8187 GGAGAACCAAGAGCT 8201

RESULT 10
ADBS9205
ID ADBS9205 standard; DNA; 15231 BP.
AC ADBS9205;
DT 04-DEC-2003 (first entry)
DE Toxicity-related gene, SEQ ID 4231.
KW Toxic; toxin; gene expression profile; hepatotoxicity; liver; drug screening; toxicity assay; ds.
XX Unidentified.
OS
PN WO2003064624-A2.
PD 07-AUG-2003.
PF 31-JAN-2003; 2003WO-US003194.
XX 31-JAN-2002; 2002US-00060087.
PR 15-MAR-2002; 2002US-0364045P.
PR 15-MAR-2002; 2002US-0364055P.

30-DEC-2002; 2002US-0436643P.
(GENE-) GENE LOGIC INC.
Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;
WPI; 2003-689530/65.

Predicting a toxic effect of a compound, useful in identifying toxicity markers in liver tissues or cells for drug screening and toxicity assays, PT comprises preparing gene expression profile of tissue or cells exposed to the compound.

Claim 1; SEQ ID NO 4231, 1156pp; English.

The present invention relates to a method for predicting a toxic effect of a compound. The method comprises preparing a gene expression profile of a tissue or cell sample exposed to the compound, and comparing the gene expression profile to a database comprising SEQ ID 1-4925, where the differential expression of the gene indicates at least one toxic effect. The method is useful for predicting at least one toxic effect of a toxic compound, predicting hepatotoxicity or the progression of a toxic effect of a compound, identifying an agent that modulates the onset or progression of a toxic response, predicting the cellular pathways that a compound modulates in a cell, and identifying an agent that modulates at least one activity of a protein. The method and compositions of the present invention using a database of genes having liver toxin-induced differential expression, are useful in identifying toxicity markers in liver tissues or cells for drug screening and toxicity assays. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 15231 BP; 3451 A; 4277 C; 4932 G; 2571 T; 0 U; 0 Other:

Query Match 10.2%; Score 50.2; DB 10; Length 15231;
Best Local Similarity 49.8%; Pred. No. 0.029;
Matches 127; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

Qy 20 CGAGTCCCAAGAGCCGAGTTTGAAGGCTGGTGCAGAAATCCCGAGAGAGGCCCA 79
Db 7947 CGAGCAGAGAGAGCCGAGTTTGAAGGCTGGTGCAGAAATCCCGAGAGAGGCCCA 8006

Qy 80 GCTGTCCCTGTGAAGCGGAGGCTGGTGCAGAAATCCCGAGAGGCTGC 139
Db 8007 GCTGTCCCTGTGAAGCGGAGGCTGGTGCAGAAATCCCGAGAGGCTGC 8066

Qy 140 TGCCGTGTGAGAGAGGCTGCAGAGCTGGTGCAGAGGCTGGTGCAGAGGCTGC 199
Db 8067 TGCCGTGTGAGAGAGGCTGCAGAGCTGGTGCAGAGGCTGGTGCAGAGGCTGC 8126

Qy 200 GGAAGAAAGTCTGTGAGCTTCATCAGAACTGGCATTCGAGAGATGAGATGATTC 259
Db 8127 GCAAGAGAACTGCAAGCTTCGAGCAGCAGCGGAGAGGAGAACTACTGAGAGA 8186

Qy 260 GGGGAGAAATGCT 274
Db 8187 GGAGAACCAAGAGCT 8201

RESULT 11
ADBS3026
ID ADBS3026 standard; DNA; 15231 BP.
AC ADBS3026;
DT 04-DEC-2003 (first entry)
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:3568.
KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker; toxicity marker; toxicity progression; drug screening; primary rat hepatocyte toxicity modelling; gene; ds.

XX		Rattus norvegicus.
OS		
XX		
PN	MO2003065993-A2.	
XX		
PD	14-AUG-2003.	
XX		
PF	04-FEB-2003; 2003MO-US003482.	
XX		
PR	04-FEB-2002; 2002US-035317P.	
PR	13-MAR-2002; 2002US-0363534P.	
PR	08-APR-2002; 2002US-0370248P.	
PR	10-APR-2002; 2002US-0371134P.	
PR	10-APR-2002; 2002US-0371135P.	
PR	10-APR-2002; 2002US-0371150P.	
PR	11-APR-2002; 2002US-0371413P.	
PR	19-APR-2002; 2002US-0373601P.	
PR	19-APR-2002; 2002US-0373602P.	
PR	22-APR-2002; 2002US-0374139P.	
PR	08-MAY-2002; 2002US-0378370P.	
PR	09-MAY-2002; 2002US-0378652P.	
PR	09-MAY-2002; 2002US-0378653P.	
PR	09-MAY-2002; 2002US-0378665P.	
PR	09-JUL-2002; 2002US-0394230P.	
PR	09-JUL-2002; 2002US-0394253P.	
PR	04-SEP-2002; 2002US-0407688P.	
PR	28-JAN-2003; 2003US-0442900P.	

PA (GENE-) GENE LOGIC INC.
XX
XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M,
PI Elashoff M,
PI
XX WPI; 2003-731472/69.
DR

PT Determining if a compound induces a toxic effect on a tissue or cell, for
PT identifying hepatotoxic compounds, comprising comparing a gene expression
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT mean values.

PS Claim 44; SEQ ID NO 3568; 874pp; English.

The present invention describes a method for determining whether a compound induces a toxic effect on a tissue or cell. The method comprises preparing a gene expression profile of a tissue or cell sample exposed to the compound, and comparing the gene expression profile to a database comprising data or information on the Tox mean and non-Tox mean value. The method is useful for predicting or identifying at least one toxic effect, particularly hepatotoxicity, of a test or unknown compound. The markers listed in the specification are useful as diagnostic or toxicity markers for the prediction or identification of the physiological state of tissue or cell sample that has been exposed to a compound, or to identify or predict the toxic effects of a compound or an agent. These may also be used as markers for monitoring toxicity progression or for drug screening. The present sequence represents a primary rat hepatocyte toxicity modelling related gene sequence from the present invention.

Sequence 15231 BP; 3451 A; 4277 C; 4932 G; 2571 T; 0 U; 0 Other;

Query Match	10.2%	Score 50.2;	DB 10;	Length 15231;
Best Local Similarity	49.8%	Pred. NO. 0.029;		
Matches 127; Conservative	0;	Mismatches 128;	Indels 0;	Gaps 0;

20 CGAGTCCCAAGAGGCCGAGTTTGAGAGCGTGTGGCAGATTCCCGAGAGAGGAGGCCA 79

80 GCTGTCCCTGGTGGAGACGCGAGGGCTGGCTGGTGTATGAGAACTCTTCTCCGAGGGTGC 139

Db 8007 GCTCAGGAGGAGCAGCAGCGGCAGCAGCAGATGGAGCAGGAAAAGCAGGAGCTGCT 8066

140 TGCCTGTGTGAGAGAGCTCAGGAGCTGACAGTCTGGCGGCTTGAGCTGCT 199

Accession	Sequence	Position
Db	8067 GGCACGATGAGAGAGGCCCGCAGGCCGACGCTGAGGACAGAGAGAGGGTGTGAGAGCCCA	8126
Qy	200 GGAAGAAAGTTGCTGAGCCTCATCAGAAATGGCATCTGCAGAGGATGCAAGTCGATTC	259
Db	8127 GCAAGAGAGAACTGAGGCGTCTTGAGAGCAGCGGCGACAGAGGAGAACTACTGGCAGA	8186
Qy	260 GGGGAGAGAAATGCT	274
Db	8187 GGAGACCAGAGCT	8201

RESULT 12
ABT42078
ID ABT42078 standard; DNA; 15231 BP.

DT 26-JUN-2003 (first entry)

DE Toxicity modelling related rat gene SEQ ID No 1780.

KM Toxic effect; gene expression profile; renal toxicity; toxicity marker database; drug screening; toxicity assay; rat; ds.

OS *Rattus norvegicus*.

PN WO200295000-A2

PD 28-NOV-2002

PF 22-MAY-2002; 2002WO-US016173.

22-MAY-2001; 2001US-0292335P.

PR 19-JUN-2001; 2001US-0298925P.

10-JUL-2001; 2001US-0303808P.

28-AUG-2001; 2001US-0315047P.

22-OCT-2001; 2001US-0330462P.

21-NOV-2001; 2001US-0331805P.

PR 19-DEC-2001; 2001US-0340873P.

PR 21-FEB-2002; 2002US-0357843P.

PR 15-MAR-2002: 2002US-0364134P.
PK 21-FEB-2002: 2002US-0351844P.

PR 08-APR-2002; 2002US-0370206P

PR 08-APR-2002; 2002US-037024/E.
PR 17-APR-2002; 2002US-0372794P

21-APR-2002; 200205-03/18/9F.
PK
XX

PA (GENE-) GENE LOGIC INC.
YY

PI Mendrick D, Porter M, Johnson K, Higgs B, Cable A, Elashoff M, XY

DR WPI; 2003-148464/14.

PT Predicting at least one toxic effect of a compound, useful for toxicity modeling, assessing existing and upcoming profiles of a drug or chemical.

cell sample exposed to the compound, and comparing the gene expression

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

XX

effect of a compound. The method comprises a gene expression profile of

CC expression profile to a database comprising at least part of the data or

THIS PAGE LEFT BLANK

TITLE Group Phase I & II Team.
REFERENCE Analysis of the mouse transcriptome based on functional annotation
JOURNAL of 60,7770 full-length cDNAs
AUTHORS Nature 420, 563-573 (2002)

6 (bases 1 to 5069)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,M.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Kono,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishii,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

COMMENT CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/
Location/Qualifiers

FEATURES
source
1..5069
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="C57BL/6J"
/db_xref="FANTOM_DB:4831426119"
/db_xref="taxon:10090"
/clone="4831426119"
/issue_type="head"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
201..2888
/note="unnamed protein product; hypothetical Spectrin
repeat containing protein (InterPro|IPR002017, evidence:
InterPro)
putative"
/codon_start=1
/protein_id="BAC26351.1"
/db_xref="GI:26325194"
/translation="MTTRGRPERPMRQGEERQRIEMWLMHSEYLLADEFYRFQK
MVALBPVELQGLKEKQWLSHAQVLLHNVDNQAVLLDRLEEGSLFSRIGDSV
DEDAQRMKAEYAVAKAQRVDLAQVADQEDQVREVDQVLLKAVEVSHSL
GNCKATKATRLSTLDIAKDPGRGESIKRLEQAVGYIONTSPIGAKISEEELM
KGVLEKRVLMKEBERGLGLQSRDCEQIQQLRAELGDFKSKQRLAQSLEETVY
KTVDELVAQWRLFSQTRALASERVDRLTQKLTTPDLQSLDSVATITP
YQSMKGNTRLNHNATRAELWQRPRLNDLQMKALQRLDITPALPLASHTPLP
QLEAALTESRLKEQLAMQLKTDLGLSIFGQERRATLLQVTSYVDLHNSLQ
RSKLSQSLVQVHDFGVAQPLNRLKLDLQARQKGLPDRPGKQVOLLRLQGLS
EGDLGCTQIEAVPLAAGNSKQKQKQDQISCDQALQRLSEDLVDRCCQNVBHCFS
HRISLQQLMTATQTLTSHQDQVRLMDASQAGTFLSREKREYQVSLQALGQ
LVWKKSSPEGATVQSELKRLKESQALRLLENMULSKNQQLQKTEVDGKQVFT
NNIPKAGFLINPDPIPRQHGAPLEGHDLPHDQQLRDFEQMLQANSKRLRIT
MRVATKDLRTREVRKQLEEARIPECQHFENLRQAGQGLQVRGRLPYDAVQV
QAOGLPDRPDRFSGADCIPESBAEAVESLSPSPMPCGTALAAVAPALSAAVP
AAGRGGAQLRPGQRLPLCAVASVQWPAHLAHMAHR"

ORIGIN
Query Match 43.0%; Score 212.2; DB 4; Length 5069;
Best Local Similarity 77.8%; Pred. NO. 3.5e-42;
Matches 256; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 7 GCCCGGGGATGCCAGTGTCCCAAGAGCGCAGATTGGAGAGGCTGGCAGAAATCCCGG 66
Db 1972 GTCTGTGGAGATCTGATGTCCTCAAGAGCTGAGCTGAGACGCTGTCTGTGAATCCCG 2031

QY 67 AGAAGAGGCGCCAGCTGTCTCTGTGTGAAGCCGACGCGCTGGCTGTGTATGAGAACTCTT 126
Db 2032 AGAAAGAGGTCACAGGTGTCTCTGTGTCAAGCACTGGGCGACGCTTGATGAAGAACTTT 2091

QY 127 CTCGGAGGCTGTCTGCGGTGTGTGACAGAGAGACTCAAGGAGCTGGCAGAGTGTGGCGG 186
Db 2092 CCCGAGAGGCGCAACCAAGTGTCCAGAGAGACTCAAGAACTGTATGAGTCTTGGCAGG 2151

QY 187 CCTTGAAGCTGTGTGAAGAAAGTGTGTGAGCTCATCAGAAATGCGCATCTGCAGAGGA 246
Db 2152 CCTGTGGCTGTCTAGAGGACATGTGATCTCATGAGAAACAGACGCTGCAGAGGA 2211

QY 247 TGGAAGTGATTGCGGGAGAAATGTGTTTTCACCAACAACTCCCAAAGTCAGGATTTTC 306
Db 2212 CAGAGGTGACACGCGGGAAGAGAGAGTGTTCACCAACAACATCCCAAAGCGCGCTTTC 2271

QY 307 TCATCAATCCCATGTGATCTTATCCGAG 335
Db 2272 TCATCAACCTCAGAGACCCCATTTCCAGG 2300

RESULT 2
AK030542
LOCUS 3080 bp mRNA linear HTC 03-APR-2004
DEFINITION Mus musculus adult male pituitary gland cDNA, RIKEN full-length
enriched library, clone:530429E21 product:hypothetical Spectrin
repeat containing protein, full insert sequence.
AK030542
AK030542.1 GI:26326536
VERSION HTCC
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1
2 Carninci,P. and Hayashizaki,Y.
3 High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
10349636
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtractions of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
PUBMED 11042159

JOURNAL
PUBMED 11076861
TITLE The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
4 Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PUBMED 11076861
TITLE The FANTOM Consortium and the RIKEN Genome Exploration Research
5 Group Phase I & II Team.
REFERENCE Analysis of the mouse transcriptome based on functional annotation
JOURNAL Nature 420, 563-573 (2002)
AUTHORS

of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 3080)
 Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,F.,
 Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
 Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
 Horii,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T.,
 Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Kono,H., Kouda,M.,
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
 Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
 Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
 Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
 Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akanishi,S.,
 Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
 Muramatsu,M. and Hayashizaki,Y.
 Direct Submission
 Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail:genome-res@gs.c.riken.jp,
 URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
 Fax:81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL:http://genome.gsc.riken.jp/
 URL:http://fantom.gsc.riken.jp/
 Location/Qualifiers
 1..3080
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM:DB:5330429E21"
 /db_xref="taxon:10090"
 /clone="5330429E21"
 /sex="male"
 /tissue_type="pituitary gland"
 /clone_id="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 70..858
 /note="unnamed protein product; hypothetical Spectrin
 repeat containing protein (InterPro|IPR002017, evidence:
 InterPro)
 putative"
 /codon_start=1
 /protein_id="BAC27012.1"
 /db_xref="GI:26326537"
 /translation="MKSSPEGATWVQEBELKIMESWQALRIEENMLSIMRNOQLR
 TEVDYGRKQVFTNNIPKAGFLINPQDPIPRQHCANPLBEHDLPEHDPQLRDEQHL
 QAEKSKLRITITKRVATAKDLRTREVLQIEALIPGQHLFENLRLRPDRDSNL
 EDLRWMLYKSKLQSGHLTESSPGLTFFQSRQKRWSPCSLLQKCRVALPLQ
 LLLLLFLFLPLPAGEEBRSCLANNPASFALMLRYNPPPT"
 ORIGIN
 Query Match 38.3%; Score 188.6; DB 4; Length 3080;
 Best Local Similarity 78.0%; Pred. No. 2,6e-36;
 Matches 227; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

225 AGAACTGTCATCTGACAGATGAGTGAAGTGAATTCGGGAGAGAAATGTTTCCACAC 284
 |||||
 Db 181 AAAAAACACACACTCTGACAGAGACAGAGTGTGACACGGGAGAAACAGAGTGTTCACCAAC 240
 |||||
 Qy 285 AACATCCCAAGTCAGATTTCTCATCAATCCCATGATCTCTATTCACAG 335
 |||||
 Db 241 AACATCCCAAGCCGCGCTTTCTCATCAACCTCAGAGACCCCATTCACAG 291
 |||||
 RESULT 3
 A0033373/c
 LOCUS
 DEFINITION
 HS_2227_A2_H11_MR_CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone plate=2227 Col=22 Row=O, genomic survey
 sequence.
 384 bp DNA linear GSS 02-JUL-1998
 A0033373
 VERSION
 A0033373.1 GI:3285491
 GSS.
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.
 1 (bases 1 to 384)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,D., Young,D., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 10449764
 CONTACT: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 TEL: (206) 616-3618
 FAX: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2227 row: O column: 22
 Class: BAC ends
 High quality sequence stop: 384.
 Location/Qualifiers
 1..384
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="Plate=2227 Col=22 Row=O"
 /sex="male"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /note="Organ: sperm; Vector: pBelBAC11; BAC clones in
 E-Coli DH10B"
 ORIGIN
 Query Match 36.0%; Score 177.6; DB 9; Length 384;
 Best Local Similarity 93.8%; Pred. No. 1e-33;
 Matches 195; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

Db 175 GGCCAGCTCCTGTTTCATTGAGCTTGC 148

RESULT 4
 LOCUS AW357996 248 bp mRNA linear EST 25-APR-2001
 DEFINITION 41821 MRC 3BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AW357996
 VERSION AW357996.1 GI:6862002
 KEYWORDS EST.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecota; Bovidae; Bovinae; Bos.

REFERENCE
 1 (bases 1 to 248)
 Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casae,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-Mckown,C.G., Pertea,G., Holt,I., Karayancheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 11282978

TITLE
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases scaled and trimmed with phred v0.960904.e. Vector identified by cross_match with the -minscore 20 and -mismatch 12 options.

JOURNAL
 PUBMED
 COMMENT
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCAGTCAGCAGC
 Plate: 21 row: 1 column: 2
 Seq primer: ATTGAGTGACACTATG.
 Location/Qualifiers
 1. 248
 /organism="Bos taurus"
 /mol_type="mRNA"
 /db_xref="taxon:9913"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 3BOV"
 /note="Vector: PCMV SPOT6; Site 1: NotI; Site 2: SalI; library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."

FEATURES
 source

ORIGIN
 Query Match 33.1%; Score 163.2; DB 1; Length 248;
 Best Local Similarity 81.5%; Pred. No. 3,9e-30;
 Matches 169; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 3 GCCGGCCGGGGATGCCAGTCCCAAGAGCCGAGTTGAGAGCGTGGTGGAGAAATTC 62
 Db 17 GCGGGCCATGGAGACCCAGTCCCGGAGAGCTGAGGTAGAGACTGCTGCTGAATTC 76
 QY 63 CCGGAGAAGAGGCCCAAGTGTCCCTGGTGGAGAGCCGAGGGCTGGTGGATGAGAAAG 122
 Db 77 CCGGAGAAGAGAACCCAGCTGCTTGAATGAAAGCGTTGGCCGGCTGGATGAGAGGGC 136
 QY 123 TCTTTCGAGAGGTGCTGCGTGGTGACAGAGAGCTGAGAGCTGACAGAGTGGTGG 182
 Db 137 TCTTTCCTGAGAGAGTCTGCTGCTGCTCAAGGAGAGCTGAGAGCTGACCGAGTATGG 196
 QY 183 CCGGAGCTTGAAGCTGTGGAAGAAGTCTGAGAGCTCATCAGAAACTGGC 234
 Db 197 AGGCGCTTGAAGAGCTGAGAGAGAGCTGCTGAGAGCTCATCAGAAACCGGC 248

RESULT 5
 LOCUS AZ971121 438 bp DNA linear GSS 27-APR-2001
 DEFINITION 2M0244E01R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
 ACCESSION AZ971121
 VERSION AZ971121.1 GI:13842348
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 438)
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Iselm,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,W., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
 Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 Unpublished (2000)

TITLE
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 1000 Std Error: 0.00
 Plate: 0244 row: E column: 01
 Seq primer: CACACAGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 438.
 Location/Qualifiers
 1. 438
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUGC2M0244E01"
 /sex="Female"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /clone_lib="Mouse 10kb plasmid UUGC2M library"
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (GI:4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
 Query Match 24.6%; Score 121.4; DB 9; Length 438;
 Best Local Similarity 69.8%; Pred. No. 1.4e-19;
 Matches 178; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

QY 208 GTCTGTGAGCTCATCAGAACTGGCATCTGCAGAGATGAGTTCGGGGAGAGA 267
 Db 114 GTCTCCCAAGTCTCATGAGAAACAGACAGCTCAGAGGACAGAGTGGACAGGGGAGA 173

QY 268 AATGTTTTCACCAACATCCCAAGTTCAGATTTCATCATCAATCCGATGATCTTA 327
 DB 174 AGAGAGTTTACCAACATCCCAAGCGGGCTTTTCATCAACCTTCAGACCCCA 233
 QY 328 TTCACGAGTCTGCAACGCTGATGCTGTCTGACGAGGCTGTGGAGAGAGGCCAGGC 387
 DB 234 TTCACGAGTCTGCAACGCTGATGCTGTCTGACGAGGCTGTGGAGAGAGGCCAGGC 292
 QY 388 CCAGGTCAAGAGTGGGTGAGGCTGTCCAGACAGGCGCCCTGCTGTGGGCAACAT 447
 DB 293 AATGTCAAGTGAAGAGGTTGGGCGCTCCAGCAGATTCCTTCTGCTGAGATTCAT 352
 QY 448 GCTCTGCTTGAGGA 462
 DB 353 GCTCAGTCTAGGA 367

RESULT 6
 CN718887 596 bp mRNA linear EST 18-MAY-2004
 LOCUS E0774B03-5 NIA Mouse four-cell-Embryo cDNA library (Long) Mus
 DEFINITION Musculus cDNA clone NIA:E0774B03 IMAGE:30903086 5', mRNA sequence.
 ACCESSION CN718887
 VERSION CN718887.1 GI:47488272
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Sharov, A.A., Piao, Y., Maiba, R., Dudekula, D.B., Qian, Y.,
 VanBuren, V., Falco, G., Martin, P.R., Stagg, C.A., Basse, U.C.,
 Wang, Y., Carter, M.G., Hametani, T., Albe, K., Akutsu, H., Sharov, L.,
 Tanaka, T.S., Kimber, W.L., Yoshikawa, T., Jaradat, S.A., Pantano, S.,
 Nagasawa, R., Boheler, K.R., Taub, D., Hodges, R.U., Longo, D.L.,
 Schlesinger, D., Keller, J., Klotz, E., Kelsey, G., Umezawa, A.,
 Vasconcelos, A.L., Rosant, J., Kunath, T., Hogan, B.L., Curci, A.,
 D'Urso, M., Kelsey, J., Hilde, W., and Ko, M.S.
 Transcriptional analysis of mouse stem cells and early embryos
 PLOS Biol. 1 (3), 410-419 (2003)
 Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igun.grc.nia.nih.gov
 Plate: E0774 Row: B Column: 03
 Seq primer: M13 Reverse
 High quality sequence stop: 596
 POLYA=No.

FEATURES
 SOURCE location/Qualifiers
 1..596
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /db_xref="taxon:10090"
 /clone="NIA:E0774B03 IMAGE:30903086"
 /issue_type="4-cell stage embryo"
 /dev_stage="4-cell"
 /lab_host="DH10B"
 /clone_id="NIA Mouse four-cell-Embryo cDNA library
 (Long)"
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site1: SalI;
 Site2: NotI; Mouse cDNA project by the Laboratory of
 Genetics, National Institute on Aging (NIA), Intramural
 Research Program, NIH (http://igun.grc.nia.nih.gov/cDNA).
 This is a long-transcript enriched cDNA library (Ref.
 Genome Res. 11: 1553-1558 (2001)). (PMD: 115441991). The
 mRNAs were extracted from a pool of 360 embryos at 4-cell
 stage. Double-stranded cDNAs were synthesized with an
 Oligo(dT) primer (Invitrogen):

5'-pGACTGTTCTAGATGCGAGGCGCCCTTTTCTTTT-3' from
 10.8ng of mRNA, treated with T4 DNA polymerase, and
 purified by ethanol-precipitation. The cDNAs were ligated
 to Loner-linker L1-SalI, purified by phenol/chloroform, and
 separated from free linkers by centrifugation 100. Then, the
 cDNAs were amplified by long-range high fidelity PCR using
 Ex Tag polymerase (Takara) with a primer SalI-S. The
 products were purified by phenol/chloroform and centrifuged
 100. The cDNAs were digested with SalI and NotI enzymes
 and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
 vector. The DH10B E. coli host was transformed with the
 ligation mixture by the standard chemical method. The
 average insert size is about 2.2kb. The library was
 constructed by Yulan Piao."

ORIGIN

Query Match 23.9%; Score 117.6; DB 7; Length 596;
 Best Local Similarity 76.6%; Pred. No. 1.3e-18;
 Matches 144; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 7 GCCCGGAGATGCCAGTCCCAAGAGCCGAGTTTAAAGGCTGTGCGAGATTCGCCG 66
 DB 409 GTCTGTGGAGTGTGAGTCCCAAGAGCTGAGTCCAGACGCTGTCTGAATCCAG 468
 QY 67 AGAAGAGGCCGAGTGTCTCCGTGTGAGAGCCGAGGCTGTGATGAGAGATCTT 126
 DB 469 AGAAGAGGTCACGAGTGTCTCCGTGTGAGAGCCGAGGCTGTGATGAGAGATCTT 528
 QY 127 CTCGAGAGGCTGTGCTGCTGTGAGAGAGCTCAGGAGCTGTGAGAGTCTGCGGG 186
 DB 529 CCCAGAGGCGGACCAATGCTCAGAGAGCTGAGAGAGTGTGAGTCTTGCAGG 588
 QY 187 CCTTGAGG 194
 DB 589 CCCTGCGG 596

RESULT 7
 BY014727 368 bp mRNA linear EST 06-DEC-2002
 LOCUS BY014727
 DEFINITION Musculus cDNA clone G730047L17 5', mRNA sequence.
 ACCESSION BY014727
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 368)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamashita, I.,
 Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
 Schonbach, C., Gojobori, T., Balderelli, R., Hill, D.P., Bult, C.,
 Hume, D.A., Quackenbush, J., Schriml, L.M., Kampin, A., Matsuda, H.,
 Bateman, S., Beisel, K.W., Blake, J.A., Brad, D., Brusic, V.,
 Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A.,
 Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T.,
 Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grilmond, S.,
 Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A.,
 Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A.,
 Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R.,
 Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
 Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G.,
 Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramchandran, S.,
 Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M.,
 Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K.,
 Sultana, R., Takekawa, Y., Taylor, M.S., Teasdale, R.D., Tomita, M.,
 Verardo, R., Wagner, L., Walsby, C., Wang, Y., Watanabe, Y.,
 Wells, C., Wilming, L.G., Wyshaw-Boris, A., Yanagisawa, M., Yang, I.,
 Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Zingales, P.,
 Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura, M.,

Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishi,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shingawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL
NATURE 420, 563-573 (2002)

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216

Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishi,Y., Itoh,M., Kawai,J., Kono,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES

source

Location/Qualifiers
1..368
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="G730047L17"
/issue_type="lung"
/cell_line="RCB-0558 LLC"
/cna="11b="RIKEN full-length enriched, lung RCB-0558 LLC cDNA"

ORIGIN

Query Match 22.2%; Score 109.2; DB 5; Length 368;
Best Local Similarity 79.6%; Pred. No. 1.5e-16;
Matches 129; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 174 GAGTCGTGGCGGCGCTTGAAGCTGTGGAAGAAAGTCTGCTCATCAGAAACTGG 233
|||
DB 4 GAGCTTTGGCAGGCGCTGCGCTGTAGAGAGAAACATGCTCATGAGAAACAG 63
234 CATTCGACAGAGTGAAGTGAATGGGGGAGAAAGTGTTCACCAACATCCCA 293
|||
DB 64 CAGCTCAGAGGACAGAGGTGACACGGGGAGAAACAGAGTGTTCACCAACATCCCA 123
294 AAGTCAGATTTCTCATCAATCCCATGAGATCTATTCCAGG 335
|||
DB 124 AAGCGCGGCTTTCTCATCAACCTCAGACGCCCATTCACAGG 165

RESULT 8
CN227139
LOCUS

CN227139 822 bp mRNA linear EST 09-APR-2004

DEFINITION
RJB088A07.ab1 Rfestsis Gallus gallus cDNA 5', mRNA sequence.

ACCESSION

CN227139

GI:46330883

VERSION

CN227139.1

KEYWORDS

EST.

SOURCE

Gallus gallus

(chicken)

ORGANISM

Gallus gallus

(chicken)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 822)

Savolainen,P., Fitzsimmons,C.J., Arvestad,L., Andersson,L. and

Lundberg,J.

EST analysis of brain and testis cDNA libraries from White Leghorn

and Red Jungle Fowl

Unpublished (2004)

Contact: Peter Savolainen

Department of Biotechnology

Royal Institute of Technology, KTH

SE-106 91 Stockholm, SWEDEN

Tel: +46 (0) 8 5537 8481

Fax: +46 (0) 8 5537 8335

Email: peter.savolainen@biotech.kth.se

Seq primer: M13 reverse primer.

Location/Qualifiers

1..822

/organism="Gallus gallus"

/mol_type="mRNA"

/strain="Red Junglefowl"

/db_xref="taxon:9031"

/sex="male"

/lab_host="Electromax DH10B (Invitrogen)"

/clone_lib="Rfestsis"

/notes="Organ: testis; Vector: pSPORT-1; Site_1: Hind III;

Site_2: EcoRI; The cDNA libraries were created with the

Superscript Plasmid System (Invitrogen)."

ORIGIN

Query Match 17.4%; Score 85.6; DB 7; Length 822;
Best Local Similarity 64.8%; Pred. No. 1.5e-10;
Matches 127; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 21 GAGTCCCAAGAGCCGAGTTTGAGAGCTGTGCGAGAAATCCCGAGAGAGGCCCG 80
|||
DB 464 GAGACAGAGGTGCGACACTTGAGAGATGGCTGAGTTCCAGATGAAGATCCAG 521
QY 81 CTGTCCCTGTGGAAGCCGAGGCTGTGCTGATGAGAAATCTTCTCCGAGGTGT 140
|||
DB 522 CTGCACCTTGTGGAAGCTCATGCGCCAGCTGTGATGAGAAATCTTCCCGAGAGACT 581
QY 141 GCGGTGTGACGAGAGAGCTCAGGAGCTGCGCAGAGCTGTGCGCGGCTTGAGGCTG 200
|||
DB 582 GCCCATGTTCAGGCGAGAGCTGATCAGCTGAAGAGTCTTGAAATCACTGAAGAAATG 641
QY 201 GAAGAAAGTCTGCTGA 216
|||
DB 642 GCACTGGTCTTCTCA 657

RESULT 9
CX917800 858 bp mRNA linear EST 07-FEB-2005
LOCUS
JGI CAN6348.fwd NIH XGC_trophe4 Xenopus tropicalis cDNA clone
DEFINITION
IMAGE:7691934 5', mRNA sequence.
ACCESSION
CX917800.1 GI:58707556
VERSION
CX917800
KEYWORDS
EST.
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Silurana.

REFERENCE
AUTHORS

Richardson,P., Lucas,S., Rokhsar,D., Deter,J.C., Ng,D.C.,

TITLE Brokstein, P. and Lindquist, E.A.
 JOURNAL DOE Joint Genome Institute Xenopus tropicalis EST project
 COMMENT Unpublished (2004)
 Other ESTs: JGI CAAN6348.rev
 Contact: Lindquist, E.A., Richardson, P.
 DOE Joint Genome Institute
 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 Tel: 925 236 5600
 Fax: 925 236 5710
 Email: cdna@jgi-psf.org

Tissue Procurement: Timothy Grammer (Richard M. Harland Laboratory,
 University of California, Berkeley:
 http://tropicalis.berkeley.edu/home)
 CDNA Library Preparation: DOE Joint Genome Institute:
 http://www.jgi.doe.gov
 DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
 Clone Distribution: I.M.A.G.E. Consortium/LNL:
 http://image.lnl.gov
 Naming Conventions: EST name is generated by the concatenation of
 the JGI Clone Id and the direction of sequencing. The suffix '.fwd'
 indicates a forward sequencing read of the insert. It does not
 necessarily reflect the orientation of the insert.
 Plate: CAAN 0065 row: 9 column: 4
 High quality sequence stop: 809.

FEATURES
 source
 1..858
 Location/Qualifiers
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="IMAGE:7691934"
 /sex="male"
 /tissue_type="Testes"
 /dev_stage="Adult"
 /lab_host="Electromax DH10B"
 /clone_id="NH_XGC_tropT64"
 /note="Vector: PCMVSPORT6; Site_1: SalI; Site_2: NotI;
 This library was made from dt primed cDNA and cloned into
 Invitrogen PCMVSPORT6 vector. The work was done at DOE
 Joint Genome Institute. Poly A RNA were primed with 5'
 GACTAGTTCATGATCCGAG CGGCGCCCTTTTCTTTT 3'. cDNA
 were ligated to SalI adapter (5' TCACACCAACCGTCCG and
 5' GCGACGCGTGG), digested with NotI, size fractionated in
 1.1% agarose gel electrophoresis and ligated into NotI and
 SalI digested PCMVSPORT6 vector."

ORIGIN

Query Match 16.9%; Score 83.2; DB 8; Length 858;
 Best Local Similarity 58.5%; Pred. No. 6e-10;
 Matches 145; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 6 GGGCCGGGAGTGGCGAGTCCCAAGAGCCGAGTTGAGAGCGCTGGGAGAAATCCCG 65
 DB 64 GGGCATTTGGGATTTTGAATCCCGAGAGAAATACGATATTTGTCTGAGCTACCA 123
 QY 66 GAGAAGAGAGCCGAGCTGTCCCTGTGTGAGACCGAGGCTGTGTGATGAGAACTCT 125
 DB 124 GAAAAGATATCCAGCTACACAGGTAGAAATCTCAAGTTTGTAGTAGAGACCTCC 183
 QY 126 TCTCCGAGAGGTGTGCGGTGTGTGAGAGAGAGCTCAGGAGCTGGCAGACTCGTGGCG 185
 DB 184 TCTCCAGAGAGAGCTGCCACATTCAGACTGAGGAGAGCTTTAATGCTCTGGGCTA 243
 QY 186 GCTTTGAGGCTGTGAGAGAAAGTGTGTGAGCCCTCAGAACTGCATCTGAGAGG 245
 DB 244 AATCGAAATCTACTGTGGGGGCACTGGCAAGCGCTTTAAGACAGAGATTCCAGGA 303
 QY 246 ATGGAAGT 253
 DB 304 CTGGCACT 311

RESULT 10
 CNSOGARZ

LOCUS CNSOGARZ 1658 bp mRNA linear HTC 05-JUL-2005
 DEFINITION Tetraodon nigroviridis full-length cDNA.
 ACCESSION CR709987
 VERSION CR709987.2 GI:56291536
 KEYWORDS HTC; cDNA; full-length; full-length cDNA; Tetraodon nigroviridis.
 SOURCE Tetraodon nigroviridis
 ORGANISM Tetraodon nigroviridis

REFERENCE

1 Jallou, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N.,
 Maucell, E., Bonnaud, L., Fischer, C., Ozouf-Costaz, C., Bernot, A.,
 Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C.,
 Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boude, N.,
 Castellano, S., Anthouard, V., Jubin, C., Castell, V., Katinka, M.,
 Vacherie, B., Blemont, C., Skalli, Z., Catolico, L., Poulain, J., De
 Bernardis, V., Cruaud, C., Duprat, S., Broctier, P., Coutanceau, J.P.,
 Gouzy, J., Parra, G., Lardier, G., Chaple, C., McKernan, K.J.,
 McEwan, P., Bosak, S., Kellis, M., Wolf, J.N., Guigo, R., Zody, M.C.,
 Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D.,
 Robinson-Rechavi, M., Laudet, V., Schachter, V., Queller, F.,
 Saurin, W., Scarpelli, C., Wincker, P., Lander, E.S., Weissbach, J.
 and Roest Crollins, H.
 Genome duplication in the teleost fish Tetraodon nigroviridis
 reveals the early vertebrate proto-karyotype
 Nature 431 (7011), 946-957 (2004)
 15496914
 2 (bases 1 to 1658)
 Genoscope.
 Direct Submission
 Submitted (24-NOV-2004) Genoscope - Centre National de Sequencage -
 2 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE
 (E-mail: segre@genoscope.cns.fr - Web: www.genoscope.cns.fr)
 On Dec 3, 2004 this sequence version replaced gi:51207896.
 The sequences are based on single pass reads.
 More information available at
 http://www.genoscope.cns.fr/tetraodon.
 Location/Qualifiers
 1..1658
 /organism="Tetraodon nigroviridis"
 /mol_type="mRNA"
 /db_xref="taxon:99883"
 /tissue_type="Eggs"

COMMENT

COMMENT
 The sequences are based on single pass reads.
 More information available at
 http://www.genoscope.cns.fr/tetraodon.

FEATURES

source
 1..1658
 /organism="Tetraodon nigroviridis"
 /mol_type="mRNA"
 /db_xref="taxon:99883"
 /tissue_type="Eggs"

ORIGIN

Query Match 11.2%; Score 55.2; DB 4; Length 1658;
 Best Local Similarity 52.6%; Pred. No. 0.0073;
 Matches 120; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 4 CCGGCCCGGGAGATGCCAGTCCCAAGAGCCGAGTTTGAAGCGCTGGGAGAAATCC 63
 DB 61 CAGGAGATGAGACCTGAGAGGCTGTGACATGAGTGAAGGCCCTGGTGAGTTTC 120
 QY 64 CGGAGAGAGAGCCGAGCTGTCCCTGTGTGAGAGCGCAGGCGCTGTGTGATGAGAACT 123
 DB 121 CAGAGAGAGAGCTCCAGCTCAGACAGATGAGAGTCCAGGCTCAGAGGCTTTGAGAGAA 180
 QY 124 CTTCCTCGAGAGGTGTGCGGTGTGTGAGAGAGAGCTCAGGAGCTGGCAGAGTCTGGG 183
 DB 181 CTTCCTGTGAGAGGTGTGAGAGTGTGAGAGAGCTGAGAACTGTGTGAGAGAGTCTGGG 240
 QY 184 GGGCTTGAGAGCTGTGAGAGAAAGTGTGTGAGAGCTTCATCAGAACT 231
 DB 241 TGGCACTCTATGACATCAGTCTTAATCTGACAGAGCTGTGAACAGCT 288

RESULT 11
 CD574999/c 695 bp mRNA linear EST 12-JUN-2003
 LOCUS UCRPT01_02ca04_b1 Pontinus trifoliata CTV-challenged cDNA library -
 DEFINITION Agi Pontinus trifoliata cDNA clone UCRPT01_02ca04, mRNA sequence.

ACCESSION CD574999
 VERSION CD574999.1 GI:31670901
 KEYWORDS EST.
 SOURCE Poncirus trifoliata
 ORGANISM Poncirus trifoliata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eucosids II; Sapindales; Rutaceae; Poncirus.
 1 (bases 1 to 695)
 ROOSE,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D.,
 Wanmaker,S., Collura,K., Feuerbacher,O., Kim,H.R., Kudrna,D.,
 Wing,R. and Yu,Y.
 Development of EST Resources and New Genetic Markers for California
 Citrus - Poncirus trifoliata CTV-challenged phloem - AGI
 JOURNAL Unpublished (2003)
 COMMENT Contact: Mikeal Roose
 Department of Botany & Plant Sciences, University of California
 Riverside, CA, 92521-0124, USA
 Tel: 9097874137
 Fax: 9097874437
 Email: mikeal.roose@ucr.edu
 Seg primer: T7.
 FEATURES
 source Location/Qualifiers
 1..695
 /organism="Poncirus trifoliata"
 /mol_type="mRNA"
 /cultivar="Pomeroy OP"
 /db_xref="taxon:37690"
 /clone="UCRPT01.02ca04"
 /issue_type="Phloem"
 /dev_stage="10 - 30 cm shoots"
 /lab_host="E. coli TJC121"
 /clone_lib="Poncirus trifoliata CTV-challenged cDNA
 library - AGI"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
 greenhouse at University of California, Riverside. The
 action was a open-pollinated (very probably selfed)
 seedling of Poncirus trifoliata cv Pomeroy that was
 selected as homozygous for the CTV resistance gene. The
 rootstock was sweet orange infected with citrus tristeza
 virus (CTV) isolate T514 over 1 year before sampling (CTV
 infects sweet orange, but not genotypes carrying the CTV
 resistance gene. Shoots 10-30 cm long were harvested in
 October 2000, and the green phloem (bark) was removed and
 frozen quickly in dry ice. Total RNA was extracted using
 TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA
 library was made, and 0.5 million primary lambda cDNA
 clones were in vivo excised to give a population of
 phuescript SK(-) phagemids. All steps to this point were
 performed in the ML Roose lab at the University of
 California, Riverside by X. Ye. Phagemids were plated,
 plasmid DNA purified, cDNA clones archived, and DNA
 sequences determined bi-directionally using an ABI3730 at
 the Arizona Genomics Institute, University of Arizona
 (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu).
 Chromatogram files were transmitted to UC Riverside (by
 Yu), then processed at UC Riverside (by Wanmaker) using
 the HarVest pipeline (<http://harvest.ucr.edu>) to remove
 vector and cloning oligo sequences and various
 contaminants, and to trim to a high quality region.
 Sequences that retained a phred 17 region of at least 100
 bases were deposited to GenBank."

ORIGIN
 Query Match 10.8%; Score 53.4; DB 6; Length 695;
 Best Local Similarity 55.0%; Pred. No. 0.017;
 Matches 105; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

11 GGGGAGTGGCAGTCCCAAGAGCGCGAGTTTGAGAGCGTGGTCAGAAATTCGCCGAGAA 70
 |||||
 416 GGGGCTTCGACTTGAGAGAGAGGTGTGTGGCGCTTGGCGGCGGTGTGCGCGGA 357

71 GGAGCCCGAGCTGCTCCCTGGTGAAGCGAGGCTGCTGTGATGAGAACTCTTCTCC 130
 |||||
 356 GGAGAGACAGTGTGAGAGAGCTGATTTGGAGGTTGTGTGCTGTGTGTTAGGA 297
 |||||
 131 GGAGGCTGCTGCGTGTGTGTCAGAGAGAGCTCAGGAGCTGCGAGTGTGCGGCTT 190
 |||||
 296 GCTGTGCTGTCGAGGTGTGAGAGAGAGATTTGGAGGTGTGTGCTACTAGAGGCGGT 237
 |||||
 191 GAGCGCTGCTGG 201
 |||||
 236 GCGGCTGTGGT 226

RESULT 12
 LOCUS CV707428/c 708 bp mRNA linear EST 03-NOV-2004
 DEFINITION UCRPT01_0009M24.f Poncirus trifoliata CTV-challenged cDNA library -
 AGI2 Poncirus trifoliata cDNA clone PT_6Ea009M24, mRNA sequence.
 CV707428
 CV707428.1 GI:55289796
 EST
 Poncirus trifoliata
 Poncirus trifoliata
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eucosids II; Sapindales; Rutaceae; Poncirus.
 1 (bases 1 to 708)
 ROOSE,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D.,
 Wanmaker,S., Kim,H.R., Kudrna,D. and Stum,D., Wissecki,M.,
 Wing,R.
 Development of EST Resources and New Genetic Markers for California
 Citrus - Poncirus trifoliata CTV-challenged phloem - AGI2
 Unpublished (2004)
 JOURNAL Contact: Mikeal Roose
 Department of Botany & Plant Sciences, University of California
 Riverside, CA, 92521-0124, USA
 Tel: 9097874137
 Fax: 9097874437
 Email: mikeal.roose@ucr.edu
 Seg primer: T7.
 FEATURES
 source Location/Qualifiers
 1..708
 /organism="Poncirus trifoliata"
 /mol_type="mRNA"
 /cultivar="Pomeroy OP"
 /db_xref="taxon:37690"
 /clone="PT_6Ea009M24"
 /issue_type="Phloem"
 /dev_stage="10 - 30 cm shoots"
 /lab_host="E. coli TJC121"
 /clone_lib="Poncirus trifoliata CTV-challenged cDNA
 library - AGI2"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
 greenhouse at University of California, Riverside. The
 action was a open-pollinated (very probably selfed)
 seedling of Poncirus trifoliata cv Pomeroy that was
 selected as homozygous for the CTV resistance gene. The
 rootstock was sweet orange infected with citrus tristeza
 virus (CTV) isolate T514 over 1 year before sampling (CTV
 infects sweet orange, but not genotypes carrying the CTV
 resistance gene). Shoots 10-30 cm long were harvested in
 October 2000, and the green phloem (bark) was removed and
 frozen quickly in dry ice. Total RNA was extracted using
 TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA
 library was made, and 0.5 million primary lambda cDNA
 clones were in vivo excised to give a population of
 phuescript SK(-) phagemids. All steps to this point were
 performed in the ML Roose lab at the University of
 California, Riverside by X. Ye. Phagemids were plated,
 plasmid DNA purified, cDNA clones archived, and DNA
 sequences determined bi-directionally using an ABI3730 at
 the Arizona Genomics Institute, University of Arizona

ACCESSION CD878716
 VERSION CD878716.1 GI:32562532
 KEYWORDS EST
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.
 1 (bases 1 to 692)
 REFERENCE
 AUTHORS Genoplatte.
 TITLE Genoplatte, a major partnership french program in plant genomics
 JOURNAL Unpublished (2003)
 COMMENT Contact: Genoplatte
 Genoplatte
 93, rue Henri Rochefort 91025 EVRY CEDEX France
 Tel: 33 1 69 47 54 00
 Fax: 33 1 69 47 54 10
 This sequence has been generated in the framework of the french
 plant genomics programme 'Genoplatte' (<http://www.genoplatte.com>
 and <http://genoplatte-info.infobiogen.fr>).
 Location/Qualifiers

FEATURES
 source
 1..692
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultivar="recital"
 /db_xref="taxon:4565"
 /clone="AZ04103H19"
 /tissue_type="root"
 /clone_1b="AZ04"

ORIGIN

Query Match 10.6%; Score 52.4; DB 6; Length 692;
 Best Local Similarity 56.3%; Pred.No.0.031; Mismatches 76; Indels 0; Gaps 0;
 Matches 98; Conservative 0;

QY	28	AAGAGCCGAGTTTGAGAGGCTGTGGCAGATTCCCGAGAGAGAGCCCACTGTCCC	87
DB	575	AGGAGGTCGATTGGCGGTGTTGTGGCAGAGTCGAGCAGAGAGAGCCGATCTGGCGG	516
QY	88	TGTTGGAAGGCGAGGCGCTGGCTGTGATGAGAGAAGTCTTCTCCGAGAGGATGCTGCCGTGG	147
DB	515	TACGAGAGTCGAGAGGAAGGTGAGCTGGCGAGCGCGGAGTCGTCGTTGATGGATTGG	456
QY	148	TGCAGAGAGAGCTCAGGAGAGCTGGCAGAGTCGTGGCGGCGCTTGAGAGGCTGCTGG	201
DB	455	TGTTGAGTCGAGATGATCTGGGCGAAGCGGTGCTGGCGGAGAGGATGGCGG	402

Search completed: April 27, 2006, 06:02:36
 Job time : 3767 sec

DEVELOPMENTAL STAGE: Somatic
TISSUE TYPE: Root
US-07-915-246-1

Query Match 10.4%; Score 51.4; DB 2; Length 1505;
Best Local Similarity 49.8%; Pred. No. 0.0005;
Matches 130; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 14 GGATGCCAGATCCCAAGAGCCGAGTTTGAGAGCTGTGCGACGAATTTCCCGAAGAGA 73
DB 852 GGTATGAGAGTGAAGAGCGGTGAAGGTGTGAGCTGGATACGAGCGGAGAGCTGGA 911

QY 74 GGGCCAGCTCTCCCTGTGTGAAGCGAGGCTGTGTGTATGAGAGAATCTTCTCCGGA 133
DB 912 GGACATGTGTGAGGTGAGAGAGCGAGAGGTGTGTGTGAGAGCTGGCGCAT 971

QY 134 GGGTGTGCGCGGTGTGAGAGAGCTCAGGAGCTGGCAGAGTCTGTGCGGCTTTGAG 193
DB 972 GGTGTGTGATACGTGTGTGAGAAAGTCTGTCTGTGAGAGAGATATGAGGTGCGGT 1031

QY 194 GCTGTGGAAGAAAGTCTGTGAGCTCATCAGAACTGSCATCTGCAGAGATGGAAT 253
DB 1032 GCAGGTGACATGAGAGCTGTGGAGCGGTGAAATGAGCGGTGAGAGGTGAGGT 1091

QY 254 GGATTCGGGGAAGAAATGCT 274
DB 1092 GCACAGGTGTGTGATACGT 1112

RESULT 2

US-08-781-891-208/C
Sequence 208, Application US/08781891
Patent No. 6090620

GENERAL INFORMATION:

APPLICANT: Fu, Ying-Hui

APPLICANT: Yu, Chang-En

APPLICANT: Oshima, Junko

APPLICANT: Mulligan, John T.

TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO

NUMBER OF SEQUENCES: 209

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/781,891

FILING DATE: 27-DEC-1996

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: No. 6090620tenburg Ph.D., Carol

REGISTRATION NUMBER: 39,317

REFERENCE/DOCKET NUMBER: 240052.419

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 208:

SEQUENCE CHARACTERISTICS:

LENGTH: 16442 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-781-891-208

Query Match 10.1%; Score 49.8; DB 3; Length 16442;
Best Local Similarity 49.4%; Pred. No. 0.0031;
Matches 129; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 9 CCGGGGATCCGAGTCCCAAGAGCCGAGTTTGAGAGCTGTGCGACGAATTTCCCGAG 68
DB 16369 CAGAGCAGAGAGAGAGCAGAGCAGAGCAGAGCAGAGAGAGAGAGAGAGAGAG 16310

QY 69 AAGAGGCCAGCTGTCTCTGTGTGAAGCGCAGGCTGTGTGTATGAGAGAATCTTTCT 128
DB 16309 CAG 16250

QY 129 CCGAGAGGCTGTCCGTGTGACGAGAGAGCTCAGGAGCTGACAGATCTGTGCGGCCC 188
DB 16249 CAG 16190

QY 189 TTGAGCTGTGGAAGAAAGTCTGTGAGCTCATCAGAACTGSCATCTGCAGAGATG 248
DB 16189 AAG 16130

QY 249 GAAGTGGATTCCGGGAGAGAA 269
DB 16129 AAGAGAGAGAGAGAGAGAGAA 16109

RESULT 3

US-09-618-166-208/C
Sequence 208, Application US/09618166
Patent No. 6583112

GENERAL INFORMATION:

APPLICANT: Fu, Ying-Hui

APPLICANT: Yu, Chang-En

APPLICANT: Oshima, Junko

APPLICANT: Mulligan, John T.

TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO

NUMBER OF SEQUENCES: 209

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed Intellectual Property Law Group

STREET: 701 Fifth Avenue, Suite 6300

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/618,166

FILING DATE: 17-Jul-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Mcmasters, David D.

REGISTRATION NUMBER: 33,963

REFERENCE/DOCKET NUMBER: 240052.419C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 208:

SEQUENCE CHARACTERISTICS:

LENGTH: 16442 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 208:

US-09-618-166-208

Query Match 10.1%; Score 49.8; DB 3; Length 16442;
Best Local Similarity 49.4%; Pred. No. 0.0031;
Matches 129; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 9 CCGGCGGATCCCGAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGGCAGAAATTCGCCGAG 68
 DB 16369 CAG 16310
 QY 69 AAGGAGGCCCACTGTCTCCCTGTGTGAAGCCGAGGCTGTGGTGTATGAGAAATCTTCT 128
 DB 16309 CAG 16250
 QY 129 CCGGAGGAGTCTGCTCCGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 188
 DB 16249 CAG 16190
 QY 189 TTGAGCTGTGTGAAGAAAGTGTGCTGAGCTCATCAGAACTGTCAGATG 248
 DB 16189 AAG 16130
 QY 249 GAAGTGAATTCGGGAGAGAA 269
 DB 16129 AAGAGAGAGAGAGAGAGAGAA 16109

RESULT 4
 US-10-148-806-3/c
 ; Sequence 3, Application US/10148806
 ; Patent No. 6762042
 ; GENERAL INFORMATION:
 ; APPLICANT: Bai, Chang
 ; APPLICANT: Metzger, Michael
 ; APPLICANT: Liu, Xiaomei
 ; TITLE OF INVENTION: DNA MOLECULES ENCODING HUMAN NHL, A DNA
 ; FILE REFERENCE: 20585P
 ; CURRENT APPLICATION NUMBER: US/10/148, 806
 ; CURRENT FILING DATE: 2002-06-05
 ; PRIOR APPLICATION NUMBER: US00/33065
 ; PRIOR FILING DATE: 2000-12-09
 ; PRIOR APPLICATION NUMBER: 60/169,970
 ; PRIOR FILING DATE: 1999-12-09
 ; NUMBER OF SEQ ID NOS: 38
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 114793
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; US-10-148-806-3

Query Match 9.9%; Score 48.6; DB 3; Length 114793;
 Best Local Similarity 54.9%; Pred. No. 0.013;
 Matches 96; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
 QY 30 GAGCGCCGAGTTGAGAGGCTGTGCAAAATTCGCCGAGAGAGAGAGAGAGAGAGAGAG 89
 DB 82835 GAGGAG 82776
 QY 90 GTGGAAGGCGAGGCTGTGCTGTATGAGAAATCTTCCGAGAGAGAGAGAGAGAGAGAG 149
 DB 82775 GTGAG 82716
 QY 150 CAGAGAGAGCTCAGAGAGAGCTGCGAGAGTGTGCGCGGCTTGAAGCTCTGTGAAG 204
 DB 82715 GAGGT 82661

RESULT 5
 US-08-217-327-3/c
 ; Sequence 3, Application US/08217327
 ; Patent No. 5474925
 ; GENERAL INFORMATION:
 ; APPLICANT: John, Maliyakal E
 ; APPLICANT: Barton, Kenneth A
 ; TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
 ; NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Quarles and Brady
 STREET: P.O. Box 2113
 CITY: Madison
 STATE: WI
 COUNTRY: USA
 ZIP: 53701-2113
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/217,327
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/812,233
 FILING DATE: 19-DEC-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J
 REGISTRATION NUMBER: 27,386
 REFERENCE/DOCKET NUMBER: 1122990831
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 608-251-5000
 TELEFAX: 608-251-9166
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 913 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Gossypium hirsutum
 STRAIN: Coker 312
 TISSUE TYPE: Fiber cells
 IMMEDIATE SOURCE:
 CLONE: H6
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 72..716
 FEATURE:
 NAME/KEY: sig_peptide
 LOCATION: 72..147
 US-08-217-327-3

Query Match 9.7%; Score 47.8; DB 2; Length 913;
 Best Local Similarity 49.4%; Pred. No. 0.0039;
 Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
 QY 23 GTCCAGAGAGCCGAGTTTGAAGAGCTGTGGCAGAAATTCGCCGAGAGAGAGAGAGAG 82
 DB 517 GTGCTGTGAGAGAGAGCAATGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 458
 QY 83 GTCCCTGTGAG 142
 DB 457 GTGTGGCTGTGAG 398
 QY 143 CGTGTGTGAG 202
 DB 397 GTGTGAG 338
 QY 203 AGAAGTCTGTGAG 262
 DB 337 GAG 278
 QY 263 GAAG 273
 DB 277 GATGAG 267

```
RESULT 6
US-07-885-970A-3/c
; Sequence 3, Application US/07885970A
; Patent No. 5495070
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; City: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,970A
; FILING DATE: 19920518
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB15A1
; CLONE: H6
; US-07-885-970A-3

Query Match          9.7%; Score 47.8; DB 2; Length 913;
Best Local Similarity 49.4%; Pred. No. 0.0039;
Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
```

```
Db          337 GGAGGAGAACTGGGTGAGAGATTGAAACTGGGGGTGAGAGAACTGTCACTGGGGGCGGT 278
Qy          263 GAAGAAATGG 273
Db          277 GATGAAGTGGG 267

RESULT 7
US-08-298-687A-3/c
; Sequence 3, Application US/08298687A
; Patent No. 5521078
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; City: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,687A
; FILING DATE:
; CLASSIFICATION: 800
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB15A1
; CLONE: H6
; US-08-298-687A-3

Query Match          9.7%; Score 47.8; DB 2; Length 913;
Best Local Similarity 49.4%; Pred. No. 0.0039;
Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
```

Db 457 GGTGGGTTGACAGTGTGGAGTGGCCGAGAGTGAAGCTGTGGAGGATGGCAGCA 398
QY 143 CGTGTGCAGAGAGAGCTCAGGAGCTGGCAGAGTCTGGCGGCTTGAAGCTGTGGA 202
Db 397 GGTGAGAGAGCTGAGGTGGAGTGGCAGAGAGAGAGAGAGTGTGGATTTGCAGGA 338
QY 203 AGAAGTCTGTGAGCCTCATCAGAACTGGCATCTGCAGAGATGGAAGTGGCGG 262
Db 337 GGAGAGAACTGGGTGAGAGAGTGAAGCTGGGGTGGAGAACTGTCACTGGGGCGGT 278
QY 263 GAAGAAATGG 273
Db 277 GATGAAGTGG 267

RESULT 8
US-08-530-797-2/c
Sequence 2, Application US/08530797
Patent No. 5597718

GENERAL INFORMATION:
APPLICANT: John, Maliyakal E.

APPLICANT: Umbeck, Paul F.

APPLICANT: Brill, Winston J.

TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS

TITLE OF INVENTION: FOR ALTERED FIBER

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles and Brady

STREET: P.O. Box 2113

CITY: MADISON

STATE: WISCONSIN

COUNTRY: U.S.A.

ZIP: 53701

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette - 3.50 inch, 800KB storage

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh

SOFTWARE: Microsoft Word 4.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/530.797

FILING DATE: 20-SEP-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/617,239

FILING DATE: 21-NOV-90

APPLICATION NUMBER: US 07/253,243

FILING DATE: 04-OCT-88

ATTORNEY/AGENT INFORMATION:

NAME: Nicholas J. Seay

REGISTRATION NUMBER: 27,386

REFERENCE/DOCKET NUMBER: 1122990245

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 913 base pairs

STRANDEDNESS: single

TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ANTI-SENSE: no

HYPOTHETICAL: no

ORGANISM: Gossypium hirsutum

STRAIN: Coker 312

DEVELOPMENTAL STAGE: 15 day old fiber cells

TISSUE TYPE: fiber cells

IMMEDIATE SOURCE:

LIBRARY: CKFB15A1

CLONE: H6

US-08-530-797-2

Best Local Similarity 49.4%; Pred. No. 0.0039;
Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 23 GTCCCAAGAGCCGAGATTGAGAGCTGTGGCAGAAATTCGCCGAGAGAGGCCAGCT 82
Db 517 GTGGCTGAGAGAGAACCAATGAGCTGGTGGGGTAGCAGAGAGAGAGTGGCTGGC 458
QY 83 GTCCCTGTGTGAAGCCGAGGCTGGCTGTGTGATGAGAAATCTTCCGAGGGTCTGC 142
Db 457 GGTGGGTTGCAAGTGTGTGAGTGGCGGAGGTGAGAAAGCTGTGTGAGAGTGGCAGCA 398
QY 143 CGTGTGCAGAGAGAGCTCAGGAGCTGGCAGAGTCTGGCGGCTTGAAGCTGTGGA 202
Db 397 GGTGAGAGAGCTGAGGTGGAGTGGCAGAGAGAGAGAGAGTGTGGATTTGCAGGA 338
QY 203 AGAAGTCTGTGAGCCTCATCAGAACTGGCATCTGCAGAGATGGAAGTGGCGG 262
Db 337 GGAGAGAACTGGGTGAGAGAGTGAAGCTGGGGTGGAGAACTGTCACTGGGGCGGT 278
QY 263 GAAGAAATGG 273
Db 277 GATGAAGTGG 267

RESULT 9
US-08-298-829-3/c
Sequence 3, Application US/08298829
Patent No. 5620882

GENERAL INFORMATION:
APPLICANT: John, Maliyakal E.

TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON

TITLE OF INVENTION: PLANTS FOR ALTERED FIBER

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nicholas J. Seay, Quarles & Brady

STREET: P.O. Box 2113, First Wisconsin Plaza

CITY: Madison

STATE: Wisconsin

COUNTRY: USA

ZIP: 53701

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Microsoft Word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/298.829

FILING DATE: 19-OCT-1994

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/885,970

FILING DATE: 18-MAY-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/617,239

FILING DATE: 21-NOV-1990

APPLICATION NUMBER: 07/253,243

FILING DATE: 04-OCT-1988

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27,386

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 283-2478

TELEFAX: (608) 251-5139

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 913 base pairs

STRANDEDNESS: double

TYPE: nucleic acid

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

Query Match 9.7%; Score 47.8; DB 2; Length 913;

```

; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKP15A1
; CLONE: H6
; US-08-298-829-3

Query Match          9.7%; Score 47.8; DB 2; Length 913;
Best Local Similarity 49.4%; Pred. No. 0.0039;
Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 23 GTCCCAAGAGCCGAGTTTGAAGAGCTGTGGCAGAAATCCCGAGAGAGGCCCACT 82
DB 517 GTGGCTGAGAGAGCAAGCAATGAGCTGTGTGGGGGTACAGAGAGAGATTGTGGC 458
QY 83 GTCCCTGTGGAAGCCGAGGCTGGCTGTGATGGAGAACTTTCTCCGAGGGGTGCTGC 142
DB 457 GTGGGGTTTCAAGTGTGAGTGTGGCGGAGGTGAGAACTGTGTGAGAGAGTGGCAGGA 398
QY 143 CGTGTGACAGAGAGAGCTCAGAGAGCTGGCAGAGTCCGTGGCGGCTTGAAGCTGTGA 202
DB 397 GTGTGAGAACTGTGAGGTGAGTTGCAGAGAGAGAGAAACAGGTGTGAGTTGCAGGA 338
QY 203 AGAAAGTCTGTGAGCTCATCAGAAACTGGCATCTGCAGAGATGGAATTCGGG 262
DB 337 GGAGGAGAACTGGTGTGAGAGATTGAACCTGGGGGTGTGAGAAAGTGTCACTGGGGCGGT 278
QY 263 GAAGAAAATGG 273
DB 277 GATGAAGTGG 267

RESULT 10
; US-08-787-335-2/c
; Sequence 2, Application US/08787335
; Patent No. 5981834
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; APPLICANT: Umbeck, Paul F.
; APPLICANT: Brill, Winston J.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
; TITLE OF INVENTION: FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Charles and Brady
; STREET: P.O BOX 2113
; STREET: FIRST WISCONSIN PLAZA
; CITY: MADISON
; STATE: WISCONSIN
; COUNTRY: U.S.A.
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.50 inch, 800kb storage
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/787,335
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/530,797
; FILING DATE:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-88
; ATTORNEY/AGENT INFORMATION:
; NAME: Nicholas J. Seay
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990245
; INFORMATION FOR SEQ ID NO: 2:
```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 913 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 15 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKP15A1
; CLONE: H6
; US-08-787-335-2

Query Match          9.7%; Score 47.8; DB 2; Length 913;
Best Local Similarity 49.4%; Pred. No. 0.0039;
Matches 124; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 23 GTCCCAAGAGCCGAGTTTGAAGAGCTGTGGCAGAAATCCCGAGAGAGGCCCACT 82
DB 517 GTGGCTGAGAGAGCAAGCAATGAGCTGTGTGGGGTGTGAGAGAGAGATTGTGGC 458
QY 83 GTCCCTGTGGAAGCCGAGGCTGGCTGTGATGGAAGTCTTCTCCGAGGGGTGCTGC 142
DB 457 GTGGGGTTTCAAGTGTGAGTGTGGCGGAGGTGAGAAAGTGTGTGAGAGATGTCAGGA 398
QY 143 CGTGTGACAGAGAGAGCTCAGAGAGCTGGCAGAGTCTGTGGCGGCTTGAAGCTGTGA 202
DB 397 GTGTGAGAACTGTGAGGTGAGTTGCAGAGAGAGAGAAACAGGTGTGAGTTGCAGGA 338
QY 203 AGAAAGTCTGTGAGCTCATCAGAAACTGGCATCTGCAGAGATGGAATTCGGG 262
DB 337 GGAGGAGAACTGGTGTGAGAGATTGAACCTGGGGGTGTGAGAAAGTGTCACTGGGGCGGT 278
QY 263 GAAGAAAATGG 273
DB 277 GATGAAGTGG 267

RESULT 11
; US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent version 3.0
; SEQ ID NO 2
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1926)
; OTHER INFORMATION: coding strand of EBNA-1 DNA
; US-09-249-585A-2

Query Match          9.7%; Score 47.8; DB 3; Length 1926;
Best Local Similarity 52.2%; Pred. No. 0.005;
Matches 106; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 2 GGGCGGCGCGGGGATGCCGAGTCCCAAGAGCCGAGTTTGAAGAGCTGTGTGCAGAAATT 61
DB 364 GGAGGGGCGAGAGAGAGAGAGGGGCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGG 423
```

QY	62	CCCGGAGAGAGAGGCCCCA	CTGCTCCCTGCTGGAA	CGCAGGGCTGGCTGGATGAGAA	121
Db	424	GCAGAGCAGAGAGAGGGG	CAAGAGCAGAGAGGGG	CGAGGCGAGGAGGGA	483
QY	122	GTCTTCCGAGAGGCTG	CCGTGATGACAGAGAGCT	CGACAGTCTGTG	181
Db	484	GGGGCAGAGGGG	CAAGAGGGGCAAGAG	GCGGCGAGAGGGGCA	543
QY	182	GCGGCTTGAAGCTT	CTGGAAG		204
Db	544	GAGGGGCAAGAGAGAGG			566

```

RESULT 12
US-09-410-399-3
; Sequence 3' Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Colter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein-Barr virus
US-09-410-399-3

```

Query Match	9.7%	Score 47.8	DB 3	Length 1926
Best Local Similarity	52.2%	Pred. No. 0.005		
Matches 106	Conservative 0	Mismatches 97	Indels 0	Gaps 0
QY	2	GGCGCCGCCCGGGGAGTCCCGAATCCCAAGAGCCCGAGTTGAGAGCGCTGCTGGCAGAATT	61	
DB	364	GGAGGGGCGAGGCGAGGAGAGAGGGGCGAGGAGAGGAGGCGAGAGGGGCGAGAGGG	423	
QY	62	CCCGAGAAAGAGCGCCCAAGCTGTCCCTGTGGAAGCGAGGGCTGGCTGTGATGAGAA	121	
DB	424	GCAGAGACAGAGGAGAGGGGCAAGACAGAGAGGAGGGGCGAGAGGGGCAAGAGAGGAGA	483	
QY	122	GTCTTCTCCGAGGGGTCTGCCGTGTGTCAGAGAGAGCTCAAGGAGCTGGCAGATCTGTC	181	
DB	484	GGGCGAGAGGGGCGAGAGGGGCGAGAGCGAGGAGGAGGAGGCGAGACAGAGAGAGGGGCA	543	
QY	182	GGCGGCCCTTGAAGGCTGCTGGGAAG	204	
DB	544	GGAGGGGCGAGACAGAGAGAGG	566	

RESULT 13
US-07-885-970A-25/c
Sequence 25, Application US/07885970A
Patent No. 5495070
GENERAL INFORMATION:
APPLICANT: John, Maliyakal E.
TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nicholas J. Seay, Charles & Brady
STREET: P.O. Box 2113, Firec Wisconsin Plaza
CITY: Madison
STATE: Wisconsin
COUNTRY: USA
ZIP: 53701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/885,970A
FILING DATE: 19920518
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/617,239
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/253,243
FILING DATE: 04-OCT-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27,386
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 283-2478
TELEFAX: (608) 251-5139
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1984 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Gossypium barbadense
STRAIN: Sea Island
IMMEDIATE SOURCE:
LIBRARY: EMBL SI
CLONE: SIH6
US-07-885-970A-25

```

Query Match	9.7%	Score 47.8	DB 2	Length 1984
Best Local Similarity	49.4%	Pred. No. 0.0051		
Matches 124	Conservative	0	Mismatches 127	Indels 0
Gaps				0
OY	23	GTCCCAAGGCCCGACGTTTGAGAGGCTGTGTGGCAGAAATTCCTCCGAGAGAGAGGCCACGT	82	
Db	766	GTGGCTGAGAGAGAACCAATGAGCTGTGTGTGGGTAGCAGAGAGAGAGATTCTGTGC	707	
OY	83	GTCCTGTGTGAAAGCGACAGGCTGTGTGTATGAGAGATCTTCTCCGAGAGGTGTGC	142	
Db	706	GGTGGCGTTGCAGGTGTGTGTGAGTGTGGCCGAGAGTGGAGAGCTGTGTGAGAGGTGCAGGA	647	
OY	143	CTGTGTGAGAGAGAGACTCAGGAGACTGGCAGAGTCTGTGTGGCGGCTTGAGGCTCTGSA	202	
Db	646	GGTGTGAGAGAGCTGTGTGTGTGTGTGCACAGAGAGAGAACGAGGTGTGTGAGTTCCAGGA	587	
OY	203	AGAAAGTCTGTGAGACCTCATCAGAACTGTGCATCTGCAGAGATGAAATGATTCGGG	262	
Db	586	GGAGAGAACTGTGTGTGTGTGTGAGAGTTGAATCTGGGGGTGTGAGAACTGTCACTGGGGGCGGT	527	
OY	263	GAGAGAAATTCG	273	
Db	526	GATGAAAGTGGG	516	

RESULT 14
US-08-298-687A-25/c
Sequence 25, Application US/08298687A
Patent No. 5521078
GENERAL INFORMATION:
APPLICANT: John, Maliyekal E.
TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nicholas J. Seay, Charles & Brady
STREET: P.O. Box 2113, First Wisconsin Plaza
CITY: Madison
STATE: Wisconsin

Thu Apr 27 17:45:33 2006

Search completed: April 27, 2006, 03:30:35
Job time : 163 secs

us-10-071-510a-16.rni

THIS PAGE LEFT BLANK

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 03:53:01 ; Search time 608 Seconds
(without alignments)
6705.269 Million cell updates/sec

Title: US-10-071-510A-16
Perfect score: 493
Sequence: 1 cggccgccccgggagatgcc.....cctgtctcattgagcctgc 493

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications NA_Main: *
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq: *
3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq: *
4: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq: *
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq: *
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq: *
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq: *
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq: *
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq: *
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	493	100.0	493	6	US-10-071-510-16 Sequence 16, App1
2	347.6	70.5	4083	6	US-10-450-763-19758 Sequence 19758, A
3	176.4	35.8	532	6	US-10-029-386-10669 Sequence 10669, A
4	174	35.3	174	6	US-10-029-386-24369 Sequence 24369, A
5	57	11.6	6302	7	US-10-437-963-93121 Sequence 93121, A
6	56.8	11.5	2282	8	US-10-479-638-7 Sequence 7, App1
7	56	11.4	6402	7	US-10-437-963-16557 Sequence 16557, A
8	53.6	10.9	6296	7	US-10-437-963-16601 Sequence 16601, A
9	53	10.8	458	7	US-10-437-963-3998 Sequence 3998, App
10	53	10.8	1023	7	US-10-437-963-56215 Sequence 56215, A
11	53	10.8	6112	7	US-10-437-963-93061 Sequence 93061, A
12	53	10.8	6417	7	US-10-437-963-16524 Sequence 16524, A
13	52.4	10.6	6337	7	US-10-437-963-16612 Sequence 16612, A
14	50.6	10.3	2139	7	US-10-437-963-69317 Sequence 69317, A
15	50.6	10.3	4701	7	US-10-437-963-69371 Sequence 69371, A
16	50.6	10.3	6972	7	US-10-437-963-16522 Sequence 16522, A
17	50.2	10.2	1612	7	US-10-437-963-102480 Sequence 102480, App
18	50.2	10.2	15231	3	US-09-917-800A-1505 Sequence 1505, App
19	50.2	10.2	15231	3	US-10-152-319A-1780 Sequence 812, App
20	49.8	10.1	768	3	US-09-938-842A-812 Sequence 812, App
21	49.8	10.1	768	3	US-09-938-842A-812 Sequence 812, App
22	49.8	10.1	2607	7	US-10-437-963-69298 Sequence 69298, A
23	49.8	10.1	2689	7	US-10-437-963-69366 Sequence 69366, A

C 24	49.8	10.1	4524	7	US-10-437-963-41480 Sequence 41480, A
C 25	49.8	10.1	16442	6	US-10-374-077-208 Sequence 208, App
C 26	49	9.9	860	7	US-10-767-701-3533 Sequence 3533, App
C 27	48.6	9.9	2853	7	US-10-437-963-69296 Sequence 69296, A
C 28	48.6	9.9	5823	7	US-10-437-963-69258 Sequence 69258, A
C 29	48.6	9.9	114793	6	US-10-148-806-3 Sequence 3, App1
C 30	48.6	9.9	114793	6	US-10-148-806-3 Sequence 3, App1
C 31	48.2	9.8	628	6	US-10-859-792-3 Sequence 22859, A
C 32	48.2	9.8	5788	7	US-10-437-963-92331 Sequence 92331, A
C 33	48.2	9.8	6332	7	US-10-437-963-16484 Sequence 16484, A
C 34	47.8	9.7	1926	6	US-10-294-804-3 Sequence 3, App1
C 35	47.8	9.7	1926	8	US-10-194-046-3 Sequence 3, App1
C 36	47.8	9.7	2820	7	US-10-437-963-76755 Sequence 76755, App
C 37	47.8	9.7	6400	5	US-10-291-230-14 Sequence 93087, A
C 38	47.8	9.7	8705	5	US-10-291-230-14 Sequence 14, App1
C 39	47.8	9.7	8705	6	US-10-291-249-14 Sequence 16, App1
C 40	47.8	9.7	8705	7	US-10-273-678-16 Sequence 4, App1
C 41	47.8	9.7	9482	9	US-10-888-961-4 Sequence 1, App1
C 42	47.8	9.7	9600	6	US-10-278-751-1 Sequence 283, App
C 43	47.8	9.7	10233	6	US-10-050-898-283 Sequence 283, App
C 44	47.8	9.7	10285	6	US-10-050-902-283 Sequence 283, App
C 45	47.8	9.7	10330	7	US-10-656-269-24 Sequence 24, App

ALIGNMENTS

RESULT 1	US-10-071-510-16	Application US/10071510
Sequence 16, App1	US-10-071-510-16	Publication No. US20030143552A1
GENERAL INFORMATION:		
APPLICANT: Clark, Edwin		
APPLICANT: Greenfield, Tallesyn		
APPLICANT: Lu, Karen		
APPLICANT: Hartmann, Lynn		
APPLICANT: Brown, Jeffrey L.		
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR THE		
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION AND THERAPY		
FILE REFERENCE: MRI-027		
CURRENT APPLICATION NUMBER: US/10/071, 510		
CURRENT FILING DATE: 2002-02-08		
PRIOR APPLICATION NUMBER: US 60/267, 276		
PRIOR FILING DATE: 2001-02-08		
NUMBER OF SEQ ID NOS: 19		
SOFTWARE: FastSeq for Windows Version 4.0		
SEQ ID NO 16		
LENGTH: 493		
TYPE: DNA		
ORGANISM: Homo sapiens		
US-10-071-510-16		
Query Match	100.0%; Score 493; DB 6; Length 493;	
Best Local Similarity	100.0%; Pred. No. 7.4e-137; Indels 0; Gaps 0;	
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY 1	CGGCGGCGCGCGGAGTCCAGATCCCAAGAGCCGAGTTGAGAGCTGTGCAGAT	60
DB 1	CGGCGGCGCGCGGAGTCCAGATCCCAAGAGCCGAGTTGAGAGCTGTGCAGAT	60
QY 61	TCGCGAGAGAGAGCCCGAGTTCCTGCTGTGAGAGCCGAGTTCCTGCTGTGAGAG	120
DB 61	TCGCGAGAGAGAGCCCGAGTTCCTGCTGTGAGAGCCGAGTTCCTGCTGTGAGAG	120
QY 121	AGTCTTCGCGAGAGTGTGCTCGTGTGAGAGAGTGTGAGAGTGTGAGAGTGTGAG	180
DB 121	AGTCTTCGCGAGAGTGTGCTCGTGTGAGAGAGTGTGAGAGTGTGAGAGTGTGAG	180
QY 181	GCGCGGCTTGTGAGAGTGTGAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTG	240
DB 181	GCGCGGCTTGTGAGAGTGTGAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTG	240

Qy	241	GGAGATGGAAGATGGATTGGGGAGAAATGGTTTCCAAACAATCCCAAATGCAG	300
Db	241	AGAGATGGAAGTGCATTTGGGGAGAAATGGTTTCCAAACAATCCCAAATGCAG	300
Qy	301	GATTTCTCATCAATCCCATGATCTCTATTCCAGCATGTGTGACGCGTGAGTCTGTCTA	360
Db	301	GATTTCTCATCAATCCCATGATCTCTATTCCAGCATGTGTGACCGCGTGAGTCTGTCTA	360
Qy	361	GCAGGGCTGTGGAGAGAGGGGCCAGGCCCCAGAGTCAAAGAGTGGGTAGGGGGTCTCAGCA	420
Db	361	GCAGGGCTGTGGAGAGAGGGGCCAGGCCCCAGAGTCAAAGAGTGGGTAGGGGGTCTCAGCA	420
Qy	421	CAGGCCCTCCCTGTCTGGGGCAACAATGCTCTGTCTGAGAGATTGGCCACGTCCTGTCT	480
Db	421	CAGGCCCTCCCTGTCTGGGGCAACAATGCTCTGTCTGAGAGATTGGCCACGTCCTGTCT	480
Qy	481	CATTGAGCCTGC	493
Db	481	CATTGAGCCTGC	493

```

RESULT 2
US-10-450-763-19758
: Sequence 19758, Application US/10450763
: Publication No. US20050196754A1
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc
: TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
: FILE REFERENCE: 790CIP3/US
: CURRENT APPLICATION NUMBER: US/10/450.763
: PRIOR FILING DATE: 2003-06-11
: PRIOR APPLICATION NUMBER: PCT/US01/08631
: PRIOR FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: 09/540,217
: PRIOR FILING DATE: 2000-03-31
: PRIOR APPLICATION NUMBER: 09/649,167
: PRIOR FILING DATE: 2000-08-23
: NUMBER OF SEQ ID NOS: 60736
: SOFTWARE: Custom
: SEQ ID NO 19758
: LENGTH: 4083
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SIMILAR
: LOCATION: (3462)..(3854)
: OTHER INFORMATION: 62% homologous to Plasmodium 10phurae hitidine-rich
: OTHER INFORMATION: protein,accession number X01469,Smith-Waterman Score=535
US-10-450-763-19758

```

	Query Match	Similarity	Score	347.6	DB 9	Length	4083			
	Best Local	Similarity	98.9%	Pred. No. 3,4e-93						
	Matches	350	Conservative	0	Mismatches	4	Indels	0	Gaps	0
Qy	2	GGCCGGCCCCGGGGGATGCCGAGTCCCAAGAGGCCGACGTTTGAGAGGCTGTGTGCGAATT	61							
Db	2796	GGGGGCCCCGGGGGAGATGCCGAGTCCCAAGAGGCCGACGTTTGAGAGGCTGTGTGCGAATT	2855							
Qy	62	CCCCGAGAAAGAGAGCCACGCTGTCCCTGTGTGAAAGCGAGGGCTGTGCTGTATGAGAA	121							
Db	2856	CCCCGAGAAAGAGAGCCACGCTGTCCCTGTGTGAAAGCGAGGGCTGTGCTGTATGAGAA	2915							
Qy	122	GTTTTCCTCCGAGAGGCTGTCCCTGTGTGCAAGAGAGCTCAGAGAGCTGGCAGATCTGTG	181							
Db	2916	GTTTTCCTCCGAGAGGCTGTCCCTGTGTGCAAGAGAGCTCAGAGAGCTGGCAGATCTGTG	2975							
Qy	182	GGGGGCTTTGAGGCTCTGTGAGAAAGATCTGTGAGCCATCATAGAAACTGTGCATCTGCA	241							
Db	2976	GGGGGCTTTGAGGCTCTGTGAGAAAGATCTGTGAGCCATCATAGAAACTGTGCATCTGCA	3035							
Qy	242	GAGGATGGAAGTGGATTCCGGGAAAGAAATGTGTTTCAACAACAATCCCAAAATCTAG	301							
Db	3036	GAGGATGGAAGTGGATTCCGGGAAAGAAATGTGTTTCAACAACAATCCCAAAATCTAG	3095							

```
Oy      302 ATTTCTCATCAATCCCATGAGATTCCTATTCCACGGCATCGTGCAGCGGTGAGTCT   355
        |||||
Db      3096 ATTTCTCATCAATCCCATGAGATTCCTATTCCACGGCATCGTGCAGCGGCAAAATCT   3145
```

```

RESULT 3
US-10-029-386-10669/c
; Sequence 10669, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GENOTYPING
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 10669
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR14.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: P10476, EVALUE 5.70e-01
; OTHER INFORMATION: NT HIT: A3277661.1, EVALUE 5.10e-01
; OTHER INFORMATION: EST_HUMAN HIT: AW842384.1, EVALUE 3.00e-03
US-10-029-386-10669

```

Query Match	35.8%	Score 176.4	DB 6	Length 532
Best Local Similarity	99.4%	Pred. No. 2.9e-42		
Matches 177	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	40	TTGAGAGCGTGTGGCAGAAATCCCGAGAGAGAGGCCACGCTGCTGTGAAGCGC	99	
Db	425	TTTAGAGGCTGTGTGGCAGAAATCCCGAGAGAGAGGCCACGCTGCTGTGAAGCGC	366	
QY	100	AGGCGTGGCTGTGTATGAGAGATCTTCTCCGAGGGTGTCTGCGTGTGACAGAGAC	159	
Db	365	AGGCGTGGCTGTGTATGAGAGATCTTCTCCGAGGGTGTCTGCGTGTGACAGAGAC	306	
QY	160	TCAGGAGAGCTGGCAGAGCTGTGTGGGGGCGTTGAAGCTGTGGAAGAAAGTGTGCTGAG	217	
Db	305	TCAGGAGAGCTGGCAGAGCTGTGTGGGGGCGTTGAAGCTGTGGAAGAAAGTGTGCTGAG	248	

```

1 RESULT 4
2 US-10-029-386-24369/c
3 Sequence 24369, Application US/10029386
4 Publication No. US20030194704A1
5 GENERAL INFORMATION:
6 APPLICANT: Penn, Sharon G.
7 APPLICANT: Rank, David R.
8 APPLICANT: Hanzel, David K.
9 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GENOTYPING
10 TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
11 FILE REFERENCE: AEMICA-X-2
12 CURRENT APPLICATION NUMBER: US/10/029,386
13 CURRENT FILING DATE: 2001-12-20
14 NUMBER OF SEQ ID NOS: 34288
15 SOFTWARE: Annamex Sequence Listing Engine vers. 1.1
16 SEQ ID NO 24369
17 LENGTH: 174
18 TYPE: DNA
19 ORGANISM: Homo sapiens
20 FEATURE:
21 OTHER INFORMATION: MAP TO CHR14.3
22 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1

```

```

; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: SWISSPROT HIT: Q99372, EVALUE 6.60e-01
; OTHER INFORMATION: EST_HUMAN HIT: B155232.1, EVALUE 5.70e-02
US-10-029-386-24369

```

Query Match	174;	35.3%;	Score	174;	DB	6;	length	174;	
Best Local Similarity		100.0%	Pred. No.	1	1e-41;				
Matches	174;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;

QY	4.	GAGGCTGAGTGGCAAGATTCCCGGAGAAGAGGCCCAAGCTGTCCCTGAGTGAAGCCCAAGG	1
Db	174	GAGGCTGAGTGGCAAAATTCCTCGAGAAAGAGGCCCAAGCTGTCCCTGAGTGAAGCCCAAGG	111
QY	104	CTGGCTGATGATGAGAAGCTTCTCCCGAGGGTGTCTGCCCTGTATGCAAGAGACTTCAAG	16
Db	114	CTGGCTGATGATGAGAAGCTTCTCCCGAGGGTGTCTGCCCTGTATGCAAGAGACTTCAAG	55
QY	164	GGAGCTGGCAGAGTCTGTGGCGGCTTTGAGGCTTGAGGAAGAAGCTTGTCTGAG	217
Db	54	GGAGCTGGCAGAGTCTGTGGCGGCTTTGAGGCTTGAGGAAGAAGAAGCTTGTCTGAG	1

```

RESULT 5
US-10-437-963-93121/c
: Sequence 93121, Application US/10437963
: Publication No. US20040123343A1
:
GENERAL INFORMATION:
:
: APPLICANT: La Rosa, Thomas J.
: APPLICANT: Kovalic, David K.
: APPLICANT: Zhou, Yihua
: APPLICANT: Cao, Yongwei
:
: APPLICANT: Wu, Wei
: APPLICANT: Boukharov, Andrey A.
:
: APPLICANT: Barbazuk, Brad
:
: APPLICANT: Li, Ping
:
: TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21 (53221) B
: CURRENT APPLICATION NUMBER: US/10/437,963
: CURRENT FILING DATE: 2003-05-14
:
: NUMBER OF SEQ ID NOS: 204966
:
: SEQ ID NO 93121
:
: LENGTH: 6302
:
: TYPE: DNA
:
: ORGANISM: Oryza sativa
:
: FEATURE:
:
: OTHER INFORMATION: Clone ID: PAT_MRT4530_91537C.1
:
: US-10-437-963-93121

```

	Query Match	Similarity	Score	DB	Length
Best Local	55.2%	11.6%	57	7	6302
Matches	111	Conservative	0	Mismatches	90
				Indels	0
				Gaps	0

	Query	DB	Score	DB	Length
1	CGCGCGCCCCGGGGATGCCAGATGCCAAGAGCGCGAATTGGAGAGCGCTGTGTGCAGAAAT	60			
5109	CGACCGAGACTTGGCCAGACTTGTGTCTTGGGAGGTGCAGCGGACATGTGTGAGCGCGAGG	5050			
61	TCCCGGAGAAAGAGGCCACAGCTGTCTCCGTGTGAGAGCGGAGGCGTGTGCTGATGATGAGA	120			
5049	AGCCCGAAGAGTGTGTGACGACGAGAGTGTGGCGGAGAGCTGTGAGAGATGTGTGACGAGG	4990			
121	AGCTTCTTCCGAGAGGTGCTGTCCGTGTGTGCAGAGAGAGCTCAGAGAGTGTGCAGATGCT	180			
4989	ACGTGCCGAGGAAACCGAGAGAGATGTGTGCAGAGAGAGATGTGCGGAGCCGAGAGATGATG	4930			
181	GGCGGCGCTTGTAGAGCTCTGG	201			
4929	TGCACGAGATGCCGCTTGTG	4909			

RESULT 6
US-10-479-638-7
; Sequence 7, Application US/10479638

```

Publication No. US20040210956A1
GENERAL INFORMATION:
APPLICANT: Don A. Roth
APPLICANT: Randolph V. Lewis
APPLICANT: The University of Wyoming
TITLE OF INVENTION: Expression of Spider Silk Proteins in Higher Plants
FILE REFERENCE: WYO. 02-0004US
CURRENT APPLICATION NUMBER: US/10/479,638
CURRENT FILING DATE: 2003-12-03
PRIOR APPLICATION NUMBER: PCT/US02/18256
PRIOR FILING DATE: 2002-06-06
PRIOR APPLICATION NUMBER: 60/296,184
PRIOR FILING DATE: 2001-06-06
NUMBER OF SEQ. ID NOS: 54
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 2282
TYPE: DNA
ORGANISM: Kukulcania hibernalis
US-10-479-638-7

```

Query Match	Similarity	Score	DB	Length
Best Local	52.5%	56.8	2282	
Matches	124	Conservative	0	Mismatches 112; Indels 0; Gaps 0;
QY	38	GTTTAGAGCGCTGTGSCAGAAATTC	CCCGAGAGAGAGGCCCACTGTTCCTCTGTGAGAC	97
DB	1156	GCTCAGCGCGCTGTCGAGCAGGTTCTTGAGAGAGAGACCAAGAGCGCTAGATCTTCG		1215
QY	98	GCAGAGCGCTGCTGTGTGATGAGAGATTTCTTCCGAGAGGTCTCCGCTGTGTGACGAGAGA		157
DB	1216	GAGGAGCTGTGCTCTGTGTCTGTGTGCCGATCCGGAATGAGAGGCTATGTGTTCTGTGATCA		1275
QY	158	GCTCAGAGAGCTGSCAGAGTCTGTGCCGAGCCTTTAGAGCTCTGSAAGAAATCTTCCTGAG		217
DB	1276	GATCTGAGCTGTGCGCGCGGTGCTGTGGGACAGCCAGATGTGTGAAGAGGTTTGTGTGAA		1335
QY	218	CCTATATGAAGAACTGCACATCTGCAGAGAGATGAGATGGATTCGGGAGAAAGAAATATG		273
DB	1336	GACCAAGGCTATGAGACAGAGTCTGTGGGACAGGCTTTGTGCTGTGGCGAGGTGTGG		1391

```

RESULT 7
US-10-437-963-16557/c
; Sequence 16557, Application US/10437963
; Publication NO. US20040123343A1
GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Bardazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 16557
; LENGTH: 6402
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_22294C.1
US-10-437-963-16557

```

Query Match	11.4%	Score 56;	DB 7;	Length 6402;
Beet Local Similarity	55.7%;	Pred. NO. 4.9e-06;		
Matches 107;	Conservative	0;	Mismatches 85;	Indels 0;
				Gaps 0;

```
QY 10 CCGGGGATGCCAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGGAGCAATTCCCGAGA 69
DB 5146 CCGAGAGAGCGGAGCTTGGAGAGATGTGCAGAGAGTGGCAGAGAGCCCGAGG 5087
QY 70 AGAGAGCCAGCTGTCTCTGTGGAAGCCGAGGCTGTGTGTATGAGAACTTTCTC 129
DB 5086 AGATGTGACAGAGAGGTGGCCGAGAGAGCCGAGAGATATGTGCAAGAGAGTGGCCG 5027
QY 130 CCGAGGCTGTCTCCCTGTGTGACAGAGAGCTCAGAGAGCTGTGTGCGGCGCT 189
DB 5026 AGAGAGCCGAGAGATGTGTGACAGAGAGAGCCGAGCGGAGAGATGTGTGACAGAGA 4967
QY 190 TGAGGCTGTG 201
DB 4966 CCGCGCTTGTG 4955
```

RESULT 8

```
US-10-437-963-16601/C
; Sequence 16601, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 2003-05-14
; SEQ ID NO 16601
; LENGTH: 6296
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_2234C.1
US-10-437-963-16601
```

```
Query Match 10.9%; Score 53.6; DB 7; Length 6296;
Best Local Similarity 54.6%; Pred. No. 2.5e-05;
Matches 107; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
```

```
QY 1 CCGCGGCGCCGGGGAGTCCGAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGTGCAAGAT 60
DB 5125 CCGTGAGCTTGGCAAGACCTTGTCTTGGAGGTGCGAGCGGACAGGTGAGAGGAGG 5066
QY 61 TCCCGAGAGAGAGAGCCAGCTGTCTCTGTGGAAGCCGAGGCTGTGTGTATGAGAGA 120
DB 5065 ARCCGAGAGAGATGTGTGAGCGGAGGTGCGAGAGAGCCGAGAGTGTGTGACAGAGG 5006
QY 121 AGTCTTCTCCGAGAGGCTGTGTGCTGTGTGCAAGAGAGCTCAGAGAGCTGTGTGCA 180
DB 5005 ACCTGTGTGAGAGAACTGTGAGAGATGTGTGACAGAGAGATGTGTGCGGAGAGATGG 4946
QY 181 GCGGCGCTGTGAGAGCT 196
DB 4945 TGCACGAATGCTCT 4930
```

RESULT 9

```
US-10-437-963-3998/C
; Sequence 3998, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
```

```
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 2003-05-14
; SEQ ID NO 3998
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_10920C.1
US-10-437-963-3998
```

```
Query Match 10.8%; Score 53; DB 7; Length 458;
Best Local Similarity 52.0%; Pred. No. 2e-05;
Matches 142; Conservative 0; Mismatches 130; Indels 1; Gaps 1;
```

```
QY 1 CCGCGGCGCCGGGGATGCCAGTCCCAAGAGCCGAGTTTGAGAGGCTGTGTGCAAGAT 60
DB 373 CCGCGGAGAGAGAGAGGCGGCGAGAGAGAGAGAGAGGCGGCGGCGGCGGAGG 314
QY 61 TCCCGAGAGAGAGCCAGCTGTCTCTGTGGAAGCCGAGGCTGTGTGTATGAGAGA 120
DB 313 AGAGAGAGAGATGTGTGCGAGAGAGAGAGAGAGAGATGTGTGCGGCGGCGGCGG 254
QY 121 AGTCTTCTCCGAGAGGCTGTGTGCTGTGTGCAAGAGAGCTCAGAGAGCTGTGTGAGAGT 179
DB 253 AGCGGAGCGGAGAGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 194
QY 180 TCGCGGCGCTTGAAGCTGTGTGGAAGAACTGTGTGAGCTGTGTATGAGAACTGGATCTG 239
DB 193 TCGCGGTGGCGCGCGCGGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 134
QY 240 CAGAGATGGAAGTGTGATTCGGGGAAGAAATG 272
DB 133 CCGAGAGAGAGATGTGTGTGTGCGGAGCAAAATTG 101
```

RESULT 10

```
US-10-437-963-56215/C
; Sequence 56215, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 2003-05-14
; SEQ ID NO 56215
; LENGTH: 1023
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_58148C.1
US-10-437-963-56215
```

```
Query Match 10.8%; Score 53; DB 7; Length 1023;
Best Local Similarity 55.0%; Pred. No. 2.4e-05;
```

Matches	104;	Conservative	0;	Mismatches	85;	Indels	0;	Gaps	0;
OY	13	GGGATGCCGAGTCCCAAGAGGCCGAGTTTGAAGAGCTGTGGCAAAATTC	CCCGAGG	72					
Db	220	GGGAGGTGCACAGCCGACACGGTGGAGGTGGAGGCCGAGAGAGCCGAGGAGAA	161						
OY	73	AGGCCCACTGTCCCTGGTGGAGCGCAGGCGTGGTGTGATGAGAAATCTTCCGG	132						
Db	160	TGCTGTCAAGAGAGGTGTGTAAGAGAGCCGAGAGATGTTGTCAGGAGGTTGGCAGAGG	101						
OY	133	AGGGTCTGCCGTGTGTGCAGAGAGGCTTCAGGAGACTGTGCAGAAATCTGTGCCGGGCTTGA	192						
Db	100	AGCCGGAAGAGATGTGTGCAGAGAGAGACGGCCGAGCAGAGAGATGTGTGCACGAGACGC	41						
OY	193	GGCTGCTGG	201						
Db	40	CGCTTGTGTG	32						

```

RESULT 11
US-10-437-963-93061/c
/ Sequence 93061, Application US/10437963
/ Publication No. US20040123343A1
/
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Wu, Wei
/ APPLICANT: Boukharov, Andrey A.
/ APPLICANT: Barbazuk, Brad
/ APPLICANT: Li, Ping
/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53221)B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ CURRENT FILING DATE: 2003-05-14
/ NUMBER OF SEQ ID NOS: 204966
/
/ SEQ ID NO 93061
/
/ LENGTH: 6112
/
/ TYPE: DNA
/ ORGANISM: Oryza sativa
/
/ FEATURE:
/
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_91482C.1
/
/ US-10-437-963-93061

```

	Query Match	Similarity	Score	DB	Length
	Best	Local	Similarity	55.8%	Pred. No. 3.8e-05;
Matches	101;	Conservative	0;	Mismatches	80;
				Indels	0;
				Gaps	0;

QY	21	GAGTCCCAAGAGCCGAGTTTATAGAGGCTGTGGCAGAAATTC	CGCGAAGAGAGGCCAG	80
DB	4855	GAGGTGCAGGCCGAGACCGGTGAGGCGGAGGCGGAGAGCC	CGAGAGATGTGTGAG	4799
QY	81	CGATCCGTGTGTAAGCCGAGGAGCTGTGCTGTATGTAAGTCTTCT	CCGAGAGGAGTCT	140
DB	4795	CAGAGAGTGGCCGAGAGGAGCCGAAAGAGATGTGTGCACAGAGGTGG	CCGAGAGGACCAAG	4736
QY	141	GCCGTGTGTCAGAGAGAGCTCAGGAGAGTCTGSCAAGATGTGTGCGGAGCCTT	TAGGCTGCTG	200
DB	4735	GAGATGTGTCAGAGAGAGATGTGCGGAGCCGAGAGAGATGTGTCCAGAAATGCCGCTT	GTTC	4676
QY	201	G	201	
DB	4675	G	4675	

RESULT 12
US-10-437-963-16524/C
; Sequence 16524, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

```

APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 16524
LENGTH: 6417
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_22264C.1
US-10-437-963-16524

Query Match          10.8%; Score 53; DB 7; Length 6417;
Best Local Similarity 55.0%; Pred. No. 3.8e-05;
Matches 104; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY      13 GGGATGCCGAGTCCCAAGAGCCGAGATTGGAAGCGTGTGGCGAATTCGCCGAGAAGG 72
DB      5115 GGGAGGTGCAGGCGGACACGCGTGTGAGGTGAGGCGGAGAGCGCGAGAGCCGAGAGGA 5056

QY      73 AGGCCAGCTGTCCCTGGTGTGAAGCCGACGGCGCTGCGTGTGAATGAGGAATCTTCTCCGG 132
DB      5055 TGGTTCAGAGAGAGGTGTGTGAGAGGCCGAGAGGAGTGTGTGCGAGAGAGTGGACAGAGG 4996

QY      133 AGGCTGTGCGCTGTGTGACGAGAGAGCTCAGGAGAGCTGCAGAGTGCAGGCGGACTTGA 192
DB      4995 AGCCGAGAGAGTGTGTGACGAGAGAGACGCGCGAGGCCGAGAGAGATGTGTGACGAGAGCC 4936

QY      193 GGTGTCTGG 201
DB      4935 CGCTTGTGTG 4927

```

```

RESULT 13
US-10-437-963-16612/c
; Sequence 16612, Application US/10437963
; Publication NO. US20040123343A1
GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 16612
; LENGTH: 6337
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_22344C.1
US-10-437-963-16612

```

Query Match	10.6%	Score 52.4;	DB 7;	Length 6337;
Best Local Similarity	54.0%	Pred. No. 5.8e-05;		
Matches 107;	Conservative	0;	Mismatches 91;	Indels 0;
				Gaps 0;

```
QY      1 CGGCGGCGCCGGGGGATGCCGAGTCCCAAGAGCGCCAGTTTGTAGAGGCTGTGCAGAAAT 60
DB      5166 CGGTGAGACTTGGCCAAACACCTTGTCTTGGAGAGTGCAGGGCGACACGCTGTGAGCCGAGG 5107
QY      61 TCCCGAGAGAAGAGCGCCCACTGTCCCTGTGTGAAGCGCAGGGCTGTGTGTATGAGA 120
DB      5106 AGCCGGAGGAGATGTGTGCACAGAGAGGTGCGGAGGAGCCGAGAGATGTGTGCAGAGG 5047
QY      121 AGCTCTTCTCCGGAGGGTGTGCTGCTGTGTGCAGAGAGCTTCAGGAGCTGTGCAGACTCGT 180
DB      5046 ACCTGTGCAGAGAACCGGAGAGATGTGTGCAGAGAGATGTGTGCAGAGAGATGTGCAGG 4987
QY      181 GCGGCGCTTGAAGCTGC 198
DB      4986 AGCCGAGAGATGTGTC 4969
```

RESULT 14

```
US-10-437-963-69317/c
; Sequence 69317, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 69317
; LENGTH: 2139
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_69998C.1
US-10-437-963-69317
```

```
Query Match      10.3%; Score 50.6; DB 7; Length 2139;
Best Local Similarity 53.2%; Pred. No. 0.00015;
Matches 107; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
```

```
QY      1 CGGCGGCGCCGGGGGATGCCGAGTCCCAAGAGCGCCAGTTTGTAGAGGCTGTGCAGAAAT 60
DB      1062 CGGTGAGACTTGGCCAAACACCTTGTCTTGGAGAGTGCAGGGCGACATGTGTGAGGTGAGG 1003
QY      61 TCCCGAGAGAAGAGCGCCCACTGTCCCTGTGTGAAGCGCAGGGCTGTGTGTATGAGA 120
DB      1002 AGCCGGAGGAGATGTGTGCACAGAGAGGTGCGGAGGAGCCGAGAGATGTGTGCAGAGTGC 943
QY      121 AGCTCTTCTCCGGAGGGTGTGCTGCTGTGTGCAGAGAGCTTCAGGAGCTGTGCAGACTCGT 180
DB      942 ACCTGTGCAGAGAACCGGAGAGATGTGTGTAGAGAGATGTGTGCAGAGCTGCAGAGATGC 883
QY      181 GCGGCGCTTGAAGCTGTGC 201
DB      882 TGCACGAGACGCCGCTTGTGC 862
```

RESULT 15

```
US-10-437-963-69371/c
; Sequence 69371, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
```

```
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 69371
; LENGTH: 4701
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_70045C.1
US-10-437-963-69371
```

```
Query Match      10.3%; Score 50.6; DB 7; Length 4701;
Best Local Similarity 53.2%; Pred. No. 0.00018;
Matches 107; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
```

```
QY      1 CGGCGGCGCCGGGGGATGCCGAGTCCCAAGAGCGCCAGTTTGTAGAGGCTGTGCAGAAAT 60
DB      3470 CGGTGAGACTTGGCCAAACACCTTGTCTTGGAGAGTGCAGGGCGACACGCTGTGAGCGGAGG 3411
QY      61 TCCCGAGAGAAGAGCGCCCACTGTCCCTGTGTGAAGCGCAGGGCTGTGTGTATGAGA 120
DB      3410 AGCCGAGAGAGATGTGTGCAGTGTGAGAGTGTGCGGAGAGCCCGAGAGATGTGTGCAGAGG 3351
QY      121 AGCTCTTCTCCGGAGGGTGTGCTGCTGTGTGCAGAGAGCTTCAGGAGCTGTGCAGACTCGT 180
DB      3350 ACCTGTGCAGAGAACCGGAGAGATGTGTGCAGAGAAATGTGCGGAGCCGAGAGAGATGC 3291
QY      181 GCGGCGCTTGAAGCTGTGC 201
DB      3290 TGCACGAGACGCCGCTTGTGC 3270
```

```
Search completed: April 27, 2006, 05:37:22
Job time : 610 secs
```


November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New).

Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

THIS PAGE LEFT BLANK